DOCUMENT RESUME

ED 148 837	TM 006 722

AUTHOR	Willingham, Warren W.; And Others
TITLE	The CAEL Validation Report.
INSTITUTION	Cooperative Assessment of Experiential Learning Project, Princeton, N.J.
SPONS AGENCY	Ford Foundation, New York, N.Y.; Fund for the Improvement of Postsecondary Education (DHEW), Washington, D.C.; Lilly Endowment, Inc., Indianapolis, Ind.
PUB DATE	Dec 76
NOTE	556p.
AVAILABLE FROM	Cooperative Assessment of Experiential Learning, American City Building, Suite 403, Columbia, Maryland 21044 ($15.00)
EDRS PRICE	MF-$1.00 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS	College Credits; College Students; Data Analysis; Data Collection; Educational Objectives; Evaluation Criteria; *Evaluation Methods; Evaluators; *Field Studies; Higher Education; Informal Assessment; Institutional Research; Interpersonal Competence; Knowledge Level; *Learning Experience; Post Secondary Education; Program Development; *Program Evaluation; Reliability; Research Methodology; Resource Materials; Special Degree Programs; *Student Evaluation; Validity; Work Experience
IDENTIFIERS	*Cooperative Assessment of Experiential Learning; *Experiential Learning; Portfolios

ABSTRACT

The Cooperative Assessment of Experiential Learning (CAEL) Project involves Educational Testing Service and some 250 member institutions. The purpose of the project is to develop procedures and materials that will be useful in strengthening the assessment of experiential learning in higher education. This report describes extensive findings that focus on the strengths and weaknesses of different evaluation methods, and other characteristics of assessment procedures that affect reliability, validity, and usefulness to students. Studies of the assessment of interpersonal skills, portfolio assessment, assessment of work competences, as well as selected institutional studies of expert evaluator judgment are described. The report provides a basic reference for institutions that wish to examine the reliability and validity of their own assessment procedures and the soundness of their underlying educational rationales. (Author/MV)

Documents acquired by ERIC include many informal unpublished materials not available from other sources. ERIC makes every effort to obtain the best copy available. Nevertheless, items of marginal reproducibility are often encountered and this affects the quality of the microfiche and hardcopy reproductions ERIC makes available via the ERIC Document Reproduction Service (EDRS). EDRS is not responsible for the quality of the original document. Reproductions supplied by EDRS are the best that can be made from the original.

ED148837

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

"PERMISSION TO REPRODUCE THIS MATERIAL IN **MICROFICHE ONLY** HAS BEEN GRANTED BY

CAEL

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

THE
CAEL
VALIDATION
REPORT

CAEL Project Report

Cooperative Assessment of
Experiential Learning

IM006 722

CAEL STEERING COMMITTEE/BOARD OF TRUSTEES

Morris T. Keeton,*# CHAIRPERSON
Provost and Vice President
Antioch College

Richard J. Allen*
Director, Division of Arts & Sciences
The Johns Hopkins University Evening College

George Ayers*# SECRETARY effective 10/76
Vice President and Dean for Academic Affairs
Metropolitan State University

Barbara A. Barbato*#
Director of Contract Center for Individualized Study
Webster College

Neal R. Berte*#
President
Birmingham Southern College

James D. Brown, Jr.*#
President
Thomas A. Edison College

Arthur W. Chickering*#
Vice President for Policy Analysis and Evaluation
Empire State College

Hortense Dixon*
Vice President for Urban Programming
Texas Southern University

John S. Duley*#
Director, Field Experience Program
Justin Morrill College
Michigan State University

Nirelle J. Galson#
Assistant Director, Independent Study
Degree Program
Syracuse University

Sheila Gordon*#
Associate Dean for Cooperative Education
LaGuardia Community College

Cyril O. Houle*
Professor of Education
University of Chicago

Winton H. Manning*
Vice President, Development
Educational Testing Service

Larraine R. Matusak#
Dean, College of Alternative Programs
University of Evansville

Jules O. Pagano,*# VICE CHAIRPERSON
Dean, Program Development and Evaluation
Florida International University

Jean M. Pennington*
Director, Continuing Education for Women
Washington University

Jane S. Permaul#
Dean, Experimental Educational Programs
University of California at Los Angeles

David H. Provost*#
State University Dean
New Program Development and Evaluation
California State University and Colleges

Gilberto de los Santos*
Dean of Instructional Development
El Paso Community College

Robert F. Sexton#
Executive Director, Office for Experiential
Education
University of Kentucky

Peter P. Smith,*# SECRETARY prior to 10/76
President
Community College of Vermont

William G. Thomas*
Director, New Dimensions
Los Angeles Community Colleges

Urban G. Whitaker#
Dean of Undergraduate Studies
San Francisco State University

---

*Member of the CAEL Steering Committee that supervised the project during the period when the work reported in this volume was carried out.

#Member of the CAEL Steering Committee and Board of Trustees that took office October 12, 1976

CAEL is an educational association of 250 institutions of higher education and other educational organizations. It was chartered by the Regents of the State of New York in September, 1976 for the purpose of fostering experiential learning and the valid and reliable assessment of its outcomes.

CAEL (Cooperative Assessment of Experiential Learning) started in March, 1974 as a research and development project involving Educational Testing Service and a group of colleges and universities. Funding was provided initially by the Carnegie Corporation of New York and subsequently by the Ford Foundation, the Lilly Endowment, and the Fund for the Improvement of Postsecondary Education.

Through June 30, 1977 inquiries concerning activities, publications, or membership should be addressed to CAEL, Educational Testing Service, Princeton, NJ 08540.

Following June 30, 1977 all correspondence should be addressed to CAEL, Antioch College, American City Building, Suite 403, Columbia, MD 21044.

$15.00 per copy

CAEL Project Report

The CAEL Validation Report

Warren W. Willingham
and
Associates

Cooperative Assessment of Experiential Learning

December 1976

Copyright © 1978 • CAEL • All rights Reserved

CONTENTS

Contributors v

Participants in CAEL Field Research vi

Acknowledgments vii

List of Tables ix

List of Figures xiii

List of Appendices xiv

Chapter I - INTRODUCTION I-1
(Warren W. Willingham)
The Rationale of CAEL Validation I-2
Overview of this Report I-10

Chapter II - REVIEW AND REVISION OF CAEL MATERIALS II-1
(Jane Porter Stutz and Warren W. Willingham)
Description of the Reports II-2
The Tryout Process II-10
Outcome of the Tryout II-13

Chapter III - OVERVIEW OF CAEL FIELD RESEARCH III-1
(Warren W. Willingham and Kurt F. Geisinger)
Purpose and Objectives III-2
Description of the Studies III-6
General Prodecure and Schedule III-12
Interpreting the Results III-15

Chapter IV - STUDIES OF ASSESSMENT OF INTERPERSONAL SKILL IV-1
(Thomas F. Donlon and Richard R. Reilly)
Study A1—To what extent can faculty differentiate various interpersonal competences in terms of their relevance to degree programs in different subject matter areas? IV-5
Study A2—To what extent do different assessors agree on the level of interpersonal competence students have acquired and to what extent is there agreement between different methods of assessment? IV-18
Study A3—What is the educational value to the student of the assessment of interpersonal skills? IV-34
Summary IV-44

Chapter V - STUDIES OF ASSESSMENT THROUGH THE PORTFOLIO V-1
(Warren W. Willingham and Joan E. Knapp)
Study B1—Do faculty agree on the amount of credit awarded for portfolios? V-3
Study B2—Do faculty agree on what constitutes college-level learning? V-15
Study B3—To what extent do faculty agree on the type of documentation required as evidence of experiential learning, and to what extent do various types of documentation prove authentic? V-22
Study B4—What is the educative value to the student in developing a portfolio and having it assessed? V-31
Summary V-39

CONTENTS

Chapter VI - STUDIES OF ASSESSMENT OF WORK COMPETENCES VI-1
(Amiel T. Sharon and Richard R. Reilly)
  Study C1—Do faculty agree on the amount of credit to be awarded for work experience? VI-3
  Study C2—To what extent can faculty agree on assigning credit for work experience to courses within a particular occupational program? VI-13
  Study C3—What is the concurrent validity of the documentation of work experience? VI-24
  Summary VI-32

Chapter VII - SELECTED INSTITUTIONAL STUDIES OF EXPERT JUDGMENT VII-1
  Introduction VII-1
  Study D1—A Reliability Study of Expert Judgment in the Assessment of Speaking Performance of Entering College Students VII-2
  (Georgine Loacker, Alverno College)
  Study D2—Development of a Competency-Based Scale for Writing VII-10
  (Ruth Churchill, Antioch College)
  Study D3—The Use of Expert Judgment in the Pre-Graduation Conference VII-31
  (Paul I. Jacobs, Thomas A. Edison College)

Chapter VIII - CONCLUSIONS AND IMPLICATIONS VIII-1
(Warren W. Willingham)
  Assessment Strengths and Weaknesses VIII-4
  Other Findings and Implications VIII-9
  Reflections on the Validation Model and Future Work VIII-15

List of References

Appendices
  1. Participants

  2. Background Information

  3. Results of Experimental Tryout

  4. Research Materials

  5. Data Analysis

  6. Background Information on Institutional Studies of Expert Judgment

CONTRIBUTORS

Ruth Churchill
Dean for Educational Evaluation
Antioch College

Thomas F. Donlon
Senior Research Psychologist
Educational Testing Service

Kurt F. Geisinger
Visiting Research Associate
Educational Testing Service

Paul I. Jacobs
Director of Test Development and Assessment
Thomas A. Edison College

Joan E. Knapp
Executive Assistant for Special Projects
Educational Testing Service

Georgine Loacker
Co-Director for the Office of Assessment Design
Alverno College

Jane Porter Stutz
Administrative Associate
Educational Testing Service

Richard R. Reilly
Research Psychologist
Educational Testing Service*

Amiel T. Sharon
Program Director
Higher Education and Career Programs/
Center for Occupational and Professional Assessment
Educational Testing Service

Warren W. Willingham
Executive Director for Program Research
Educational Testing Service

---

*Richard R. Reilly has since left ETS and is now at American Telephone and Telegraph Company.

PARTICIPANTS IN CAEL FIELD RESEARCH

Institutions/Representatives

Alverno College
*Georgine Loacker

Antioch College
*Ruth Churchill
Robert Lewis

Community College of Vermont
*Deborah Chaffee
Larry Daloz

Delaware County Community College
*Eugene Kray
Lori Hultgren

Empire State College
*Albert Serling

Governors State University
*Betty Stanley
Robert Press

Hartwick College
*Edith Daly
Diana Christopulos
David Jeffres

LaGuardia Community College
*Sheila Gordon
Martin Cohen
Irwin Feifer
Dennis Kelly
Helen McCabe

Macalester College
*Jack Rossman
Charles Green

Metropolitan College, St. Louis University
*Virginia Beard
Dorothy Fenwick
Katie Hotz

Metropolitan State University
*James Deegan
Janet Haberg

New College, University of Alabama
*Bernard Sloan
Lisa Cain
Bev Holaday
Harriet Cabell

Institutions/Representatives

North Carolina A & T State University
*Richard Fields

Notre Dame College
*Carol Ann Collins

San Francisco State University
*Edith Arrick
Paul Breen
Urban Whitaker

Staten Island Community College
*Roslyn Attinson
Gilbert Benjamin
Vincent Massaro

Stephens College
*Lou Ann Tanner
James Waddell

Texas Christian University
*Larry D. Lauer

Thomas A. Edison College
*Paul Jacobs

University of Akron
*Elizabeth Hittle
Joyce Sullivan
Kathryn Vegso

University of California at Los Angeles
*Jane Permaul

University of Missouri-Columbia
*John Mowrer

University of Oregon
*Norman Sundberg
Richard Fehnel
Lila McQueen
Janet Moursund

Wichita State University
*Martin Reif
Galan Janeksela

---

*Site Coordinator

ACKNOWLEDGMENTS

A large number of people participated actively in this work and contributed greatly to various aspects of the research reported here or the final form of the report itself. Some are listed formally as contributors because they played central roles in carrying out the work and wrote parts of the report, but it is important to acknowledge the contribution of many others.

First, appreciation is expressed especially to the students and faculty at the eighty-odd institutions that reviewed CAEL materials and those at the twenty-four research sites who gave generously of their time and their expertise in supplying the basic data for the field research. Those institutional representatives listed on page v deserve special thanks for their energy and diligence in carrying out that data collection.

Numerous consultants contributed to this work. In the summer of 1975 the External Selection Committee (Leland Medsker, William O'Connell, and Patricia Thrash) made important selections among applicants for special CAEL research and developmental work. Irene Casper, Keitha Donnelly, Paul Jacobs, and Martha Sachs rated many student papers for two studies. Many others at institutions and on the ETS staff reviewed portions of this report and made a number of helpful suggestions. Appreciation is expressed to all of these individuals.

Appreciation is also due several ETS staff whose names do not otherwise appear here but who made important contributions to this work. Gay Marshall was helpful in various aspects of preparing the report; Robert Patrick was primarily responsible for computer programming. Linda Cokos played an important coordinating role in this and many aspects of CAEL from the start. Special thanks go to Lorraine Simon who took major responsibility for organizing the data processing and for production of this report.

Warren W. Willingham

List of Tables

		Page
Table II-1	Number of Forms Received and Percentage of Respondents Who Judged the Reports Useful	II-14
Table III-1	Hypothetical Illustrations of Analysis of Differences Across Groups: Proportion of Students in Four Institutions Judged to Have Attained the Highest Level of a Particular Competence, as Assessed by Two Methods	III-18
Table III-2	Relationship Between the Reliability of a Composite Rating and the Number of Ratings on Which the Composite is Based	III-26
Table III-3	Improvement in the Reliability of Essay Grading Through Use of Multiple Readers and Multiple Samples of the Student's Writing (From Godshalk, et al., 1966)	III-27
Table A1-1	Credit Awarded for Experientially Acquired Interpersonal Competence Relevant to Psychology Requirements in One Institution--Five Illustrative Cases Rated by Three Faculty Members	IV-9
Table A1-2	Average Amount of Credit (Semester Hours) Awarded Illustrative Cases for Interpersonal Competency Gained Through Prior Experiential Learning	IV-12
Table A1-3	Percentage of Paired Faculty Judgments Differing by Specific Numbers of Credits for Each Field and Overall	IV-15
Table A1-4	Distributions of Reliability Coefficients by Field for the Average Faculty Rating in Awarding Credit for the Illustrative Cases	IV-17
Table A2-1	Overall Ratings of Counseling Competence for 15 Students in a Single Institution Assessed by Judges Using Two Different Modes of Assessment	IV-23
Table A2-2	Average Overall Ratings of Effectiveness as a Counselor for Students at Each of Nine Institutions	IV-25
Table A2-3	Interpersonal Effectiveness in a Counseling Situation--Relationships Between Sums of Overall Ratings Derived From Two Methods of Assessment	IV-27
Table A2-4	Average Rating of Students in Analyses Samples Based on Two Methods of Assessment for Seven Different Counseling Skills	IV-28
Table A2-5	Correlations Between Summed Ratings Based on Interview and Role-Playing for Seven Traits for Available Data	IV-30
Table A2-6	Reliability Coefficients Showing the Degree of Faculty Agreement in Judging the Counseling Skill of Students--by Two Methods of Assessment in Eight Institutions	IV-31

List of Tables (Continued)

		Page
Table A3-1	Summed Ratings of Pre-Assessment and Post-Assessment Students at One Institution	IV-38
Table A3-2	Average Summed Ratings for Pre-Assessment and Post-Assessment Students at Each of Eight Institutions	IV-41
Table A3-3	Average Summed Ratings for Pre-Assessment and Post-Assessment Students at Six Institutions on Particular Questions and Overall	IV-42
Table B1-1	Sample Data Showing Total Credits Awarded in Two Independent Evaluations of 14 Illustrative Portfolios in One Institution	V-7
Table B1-2	Average Credits Awarded and Reliability of Faculty Judgments for Illustrative (Common Across Institutions) and Local Institutional Portfolios	V-9
Table B1-3	Average Credits Awarded for Illustrative Portfolios by Each of Six Institutions	V-11
Table B1-4	Frequency Distribution of Absolute Differences Between Two Credit Awards for the Same Portfolio	V-13
Table B2-1	Sample Data From One Institution Showing Five Faculty Ratings as to Whether Specific Instances of Learning Deserve College Credit (0 - Not Creditable; 1 - Lower Division Credit; 2 - Upper Division Credit; 3 - Graduate Credit)	V-17
Table B2-2	Average Rating of the Level at Which Specific Instances of Learning are Judged Creditable and the Degree of Interrater Agreement in Those Ratings (Rating of 15 Instances at 10 Institutions for Which: 0 - No Credit; 1 - Lower Division Credit; 2 - Upper Division Credit; 3 - Graduate Credit)	V-19
Table B2-3	Fifteen Learning Instances and the Average Rating Regarding the Creditability of Each (0 - No Credit; 1 - Lower Division Credit; 2 - Upper Division Credit; 3 - Graduate Credit)	V-20
Table B3-1	Sample Data From One Institution Showing Whether Each of Five Faculty Raters Judged Some Type of Letter to be Required Documentation for Each of 15 Examples of Prior Learning	V-24
Table B3-2	Proportion of All Faculty Raters (N=50) Who Indicated that Each of Four Types of Documentation was Required for Each of 15 Examples of Prior Learning	V-26

List of Tables (Continued)

Page

Table B3-3 Within Each Institution, the Relative Frequency With Which Each of Four Types of Documentation Was Required V-27

Table B3-4 The Extent of Faculty Agreement as to Whether Different Types of Documentation Should be Required for Assessment V-29

Table B4-1 Average Rating of Papers for Independent Groups of Students at Each Institution Who Had and Had Not Participated in Portfolio Assessment V-34

Table B4-2 Percentage of "High" Ratings in the Pre- and Post-Assessment Groups for Each of Five Questions V-36

Table B4-3 Pattern of Average Ratings on Specific Questions for Students Who Had Gone Through Portfolio Assessment V-38

Table C1-1 Number of Students Awarded Credit for Prior Work Experience by Two Faculty Assessors at Different Institutions in Three Occupational Subjects VI-4

Table C1-2 Average Number of Credits Awarded Students for Prior Work Experience by Institution and Occupational Field VI-7

Table C1-3 Illustrative Data Showing Number of Credits Awarded for Prior Work Experience in Law Enforcement by Two Faculty Assessors (A and B) to Eight Students at One Institution VI-8

Table C1-4 Interrater Agreement and Consistency Coefficients Indicating the Reliability of Awarding Credit for Prior Work Experience by Institution and Occupational Field VI-10

Table C1-5 Number and Percentage of Faculty Assessors Differing by a Specific Number of Credits Awarded in Four Areas of Prior Work Experience VI-12

Table C2-1 Number of Faculty Assessors Awarding Credit to Each of Five Illustrative Cases by Institution and Occupational Field VI-14

Table C2-2 Illustrative Data Showing Total Credit Awarded for Prior Work Experience in Data Processing in One Institution (Five Illustrative Cases Rated by Each Faculty Assessor) VI-16

Table C2-3 Average Amount of Credit Awarded to Illustrative Cases for Prior Work Experience in Three Occupational Fields VI-18

Table C2-4 Interrater Agreement and Consistency Found in 15 Different Departments Showing Whether Faculty Tend to Agree on Amount of Credit to be Awarded to Illustrative Cases VI-20

List of Tables (Continued)

Page

Table C2-5 Number and Percentage of Faculty Assessors' Judgments Differing by Specific Numbers of Credits Awarded to Illustrative Cases for Prior Work Experience by Occupational Field VI-21

Table C2-6 Results of Analyses Showing Whether Faculty Assessors Tend to Agree on the Breakdown of Credit Awarded in Different Content Categories Within an Occupational Subject Matter VI-23

Table C3-1 Number of Student-Supervisor Pairs Who Completed Task Inventory Checklists by Institution and Occupational Field VI-25

Table C3-2 Average Number of Checklist Tasks Checked by Students and Their Work Supervisors by Institution and Occupational Field VI-27

Table C3-3 Total Number of Tasks Checked by Students and Percentage of Those Tasks Which Are Verified by the Supervisors in Different Occupational Fields and Content Categories VI-30

Table D1-1 Number of Strengths and/or Weaknesses Agreed Upon By Both Trained Judges for Each Speaker VII-6

Table D1-2 Correlations and Reliability Coefficients for the Criteria VII-8

Table D2-1 Ratings (Average of Two Readers) of Essays of Fall 1975 Entrants VII-15

Table D2-2 Consistency of Ratings: Agreement Between Readers VII-17

Table D2-3 Consistency of Ratings: Correlations Between Readers VII-19

Table D2-4 Distribution of Initial and Course Ratings for Students Rated By at Least Two Initial Readers and Two Course Instructors VII-22

Table D2-5 Consistency of Ratings: Agreement Between Initial Reader and Course Instructor for 57 Students Originally Rated 5.0 or Above VII-24

Table D2-6 Consistency of Rating: Agreement Between Initial Readers (Average) and Instructors in 5 Courses for Students Initially Rated Competent (5 and Above) VII-26

Table D2-7 Distribution of Differences Between Average Ratings of Two or More Readers of Initial Essays and Average Ratings of Two or More Instructors at End of Courses VII-28

List of Tables (Continued)

Page

Table D3-1 Number of Conferences and Assessors in Three Academic Areas VII-34

Table D3-2 A Comparison of Actual Conference Judgments With Tape Listeners' Judgments VII-38

Table D3-3 A Comparison of Judgments by Area VII-40

Table D3-4 A Comparison of Judgments in "Old" and "New" Application Procedures VII-42

List of Figures

Figure III-1 Hypothetical Correlation Scatterplot and Corresponding Four-Fold Table Showing a Correlation of About .60 III-21

Figure III-2 Illustration of Correlation Coefficients Ranging From No Relationship (.00) to a Perfect Relationship (1.00)--Each Based Upon Two Assessments of 200 Students) III-23

Figure VI-1 Average Number of Checklist Tasks Checked by Students and Their Supervisors at Five Institutions by Content Category in Three Occupational Fields VI-29

List of Appendices

Page

APPENDIX 1 - Participants
- 1a. Implementation Committee 1-1
- 1b. Resource Panel 1-2
- 1c. Field Research Participants 1-3
- 1d. Institutions Involved in Tryout of Materials 1-4

APPENDIX 2 - Background Information
- 2a. Prospectus 2-1
- 2b. Invitation to Participate in CAEL Activities and Tryout of Materials 2-11
- 2c. Schedule for CAEL Validation Work 2-27
- 2d. Initial Communication with Participants 2-29
- 2e. General Instructions to Coordinators 2-32
- 2f. Agenda for January 14 Meeting with Site Coordinators 2-35
- 2g. Request for Feedback on Tryout of Materials 2-37
- 2h. Request for Reactions to Tentative Results 2-43
- 2i. Working Papers and Institutional Reports (1975) 2-46
- 2j. Revised Reports (1976) 2-48

APPENDIX 3 - Results of Experimental Tryout
- 3a. Critical Issues 3-1
- 3b. Content Analysis of Evaluation Forms 3-11

APPENDIX 4 - Research Materials
- 4a. Guidelines for Site Coordinators (Area A) 4-1
- 4b. General Faculty Guidelines for Interpersonal Skills CAEL Validation 4-7
- 4c. Illustrative Cases for Study A1: Management, Education, and Psychology 4-9
- 4d. Faculty Rating Forms for Study A1: Management, Education, and Psychology 4-21
- 4e. Structured Interview Guidelines for Study A2 4-24
- 4f. Role-Playing Exercise for Study A2 4-27
- 4g. Faculty Rating Form for Study A2 4-28
- 4h. Student Questionnaire for Study A3 4-29
- 4i. Guidelines for Site Coordinators (Area B) 4-30
- 4j. Illustrative Portfolio for Study B1 4-36
- 4k. Faculty Portfolio Evaluation Report for Study B1-Illustrative Portfolios 4-50
- 4l. Institutional Portfolio Summary Sheet for Study B1-Actual Local Portfolios 4-51
- 4m. Fifteen Illustrative Examples of Prior Learning Outcomes for Study B2 4-52
- 4n. Faculty Recording Form for Study B2 4-64
- 4o. Fifteen Illustrative Examples of Prior Learning Outcomes for Study B3 4-66
- 4p. Checklist for Recording Type of Documentation Required for Study B3 4-77
- 4q. Form for Recording Verification of Documentation (Guidelines for Determining Document Authenticity) for Study B3 4-78
- 4r. Student Questionnaire for Study B4 4-80
- 4s. Site Coordinator Questionnaire for Study B4 4-81

List of Appendices (Continued)

Page

APPENDIX 4 - Research Materials (Continued)
- 4t. Guidelines for Site Coordinators (Area C) 4-83
- 4u. Work Experience Evaluation Form for Study C1: Data Processing, Law Enforcement, Secretarial Science, and Management 4-89
- 4v. Illustrative Case for Study C2 4-93
- 4w. Faculty Rating Form for Study C2: Data Processing, Law Enforcement, and Management 4-100
- 4x. Student Checklist for Study C3: Law Enforcement (with Faculty Instructions) 4-103
- 4y. Guidelines and Procedures (Area D) 4-120

APPENDIX 5 - Data Analysis
- Data Analysis 5-1
- 5a. Study A1 5-5
- 5b. Study A2 5-9
- 5c. Study A3 5-16
- 5d. Study B1 5-18
- 5e. Study B2 5-23
- 5f. Study B3 5-26
- 5g. Study B4 5-30
- 5h. Study C1 5-36
- 5i. Study C2 5-41
- 5j. Study C3 5-48

APPENDIX 6 - Background Information on Institutional Studies of Expert Judgment
- 6a. Alverno College. Speaking Assessment - Instructions to Students 6-1
- 6b. Alverno College. Excerpt from Procedures Manual 6-2
- 6c. Alverno College. Excerpt from Assessment Manual 6-5
- 6d. Alverno College. Speaking Assessment - Individual Faculty Rating Form 6-9
- 6e. Alverno College. Speaking Assessment - Team Rating Form 6-10
- 6f. Alverno College. Speaking Assessment - Student Self-Rating Form 6-11
- 6g. Antioch College. English Skills Essay Grading Sheet 6-12
- 6h. Antioch College. Memo to Freshmen from Testing Office 6-13
- 6i. Antioch College. Scale for Writing Competence 6-14
- 6j. Antioch College. Memo to Faculty 6-15
- 6k. Antioch College. Writing Competency Questionnaire 6-16
- 6l. Thomas A. Edison College. The Pre-Graduation Conference: Information of Faculty Interviewers 6-17
- 6m. Thomas A. Edison College. The Pre-Graduation Conference: Information for Students 6-22
- 6n. Thomas A. Edison College. Application for Pre-Graduation Conference 6-26

Chapter I

INTRODUCTION

Warren W. Willingham

CAEL is a cooperative undertaking that involves Educational Testing Service and an Assembly of some 250 member institutions. The purpose of the project is to develop procedures and materials that will be useful in strengthening the assessment of experiential learning in higher education. Two publications provide useful background information: CAEL Project Status Report (Willingham, 1976) gives a brief overview of CAEL's origins and its various activities; Experiential Learning: Rationale, Characteristics, and Assessment (Keeton and Associates, 1976) includes a series of commissioned papers on the nature of experiential learning and problems concerning its assessment.

The CAEL project was organized in three overlapping stages: a development stage to produce new assessment procedures and materials; a validation stage to test and improve the products of such work; and a utilization stage to begin as early as possible to make this work available and to facilitate mutual problem-solving among professionals concerned with the assessment of experiential learning. The purpose of this report is to document fully the procedure and results of the validation stage of the CAEL project. This initial chapter provides a brief orientation and overview of the contents of this volume.

The Rationale of CAEL Validation

The rationale for the validation stage stems from the primary goals of the CAEL project and also the context in which it was initiated. Following the Commission on Non-Traditional Study (CNS, 1973), the critical need for improved means of assessing experiential learning was widely recognized. For several reasons the assessment problem is more complicated than is true in the case of classroom learning: experiential learning tends to be highly individualistic; by its nature the potential range of relevant learning experiences is broad and not well defined; and experiential learning is often not subject to close faculty supervision. Thus, the need for sound new methods of assessment is associated with such familiar problems as establishing credibility of programs, minimizing abuse of flexible credit policies, and insuring public recognition of credit granted. At the same time, there is the equally important problem of improving the value of assessment to the individual.

To meet such goals successfully, developmental work must clearly focus upon practices and methods which, under scrutiny, prove to be most useful and also defensible. From the very beginning those involved in the CAEL project, and others observing it, worried about "standardization." There were several aspects of this concern and each had its own legitimacy.

First, there was the fear that some might perceive the assessment of experiential learning as an overly simple problem based upon a narrow view of education. Given that an improved sense of order in this field would be desirable, there was considerable nervousness that a large effort might have the effect of imposing one educational philosophy or rationale when many abounded. Consequently, it was recognized that CAEL must involve active participation of many institutions.

A second fear was that CAEL would promote an overly simple solution, especially a traditional psychometric solution in the form of an objective examination. The need for useful forms, documents, instruments, etc. was certainly acknowledged. But it was equally clear that "helpful devices" must not have the effect of forcing learning into an administrative or psychometric solution that is expedient but perhaps procrustean and hence not useful in the long run. Consequently, it was early recognized that CAEL must not seek to design standard tests or norm-referenced techniques that might seemingly be widely applicable. The project concentrated instead upon the assessment process--the important principles that need to be understood and applied, the procedural steps that are normally useful, and prototype materials that could be adapted to particular situations and types of learning.

A third concern had to do with timing. Many educators reported a need for immediate help on assessment problems and felt that whatever CAEL could produce would likely prove useful. Thus, the danger of premature "routine" solutions was very much on the minds of the CAEL Steering Committee in the early stages of the project. As a result, CAEL initially produced working papers clearly described as "preliminary documents" to be reviewed and revised.

From these various considerations CAEL developed the following general strategy for working toward its long-range goals of improved assessment methods. During the first year, developmental work in selected areas led to a series of working papers on different assessment problems. These working papers emphasized the principles of assessment and how they might be applied to different programs. The validation stage initiated in the second year involved complementary efforts of two types--one was a broadly based trial of the working papers leading to their eventual revision; another was field research that would test principles and practices and lead to their improvement.

The Developmental Work

In its first year CAEL recognized that it could not possibly cover all areas needing developmental work; nevertheless, it was necessary to start somewhere, and CAEL selected four priority areas as follows. Noting that there are various types of competences, different arenas for learning, alternate methods of assessment, and different parties to the assessment, CAEL chose one important instance of each--

- Assessing the Achievement of Interpersonal Skills (an important competence)--Experiential learning frequently involves application of knowledge in situations that depend upon interpersonal skills and offer special advantages for developing such competences.
- Use of Portfolios in Assessing Non-Sponsored Learning (an important method)--Assessment of non-sponsored learning most often begins with preparation of a portfolio which presents pertinent information and documentation, and this process is of special current interest to many institutions.
- Assessing the Learning Outcomes of Work Experience (an important arena)--The relation of education to work has attained national significance, and the various types of "work" experience are among the most common forms of experiential learning offered for assessment.
- Use of Expert Judgment in Assessing Learning Outcomes (an important assessor)--Much experiential learning is highly individualized, based on unique combinations of experience, and must necessarily be assessed through some form of expert judgment rather than objective instruments.

During the academic year of 1974-75, ETS and the original ten CAEL task force institutions collaborated on developmental work in these four priority areas. The task force institutions included:

Antioch College
California State University and Colleges (San Francisco State University)
Community College of Vermont
El Paso Community College
Empire State College
Florida International University
Massachusetts State College System (Framingham State College)
Metropolitan State University
New College (University of Alabama)
Thomas A. Edison College

An Implementation Committee was constituted of representatives from each of these task force institutions and ETS (see Appendix 1a). This group, plus other collaborators, worked on seven developmental tasks and produced by the end of the academic year seven working papers. These working papers formed the nucleus of CAEL work during that developmental stage. They are listed in Appendix 2i and described briefly in Chapter II. At the end of the year the CAEL Steering Committee appointed a CAEL Resource Panel consisting of a number of institutional representatives who made important contributions to the work (see Appendix 1b).

Subsequently, five special projects were selected for publication as institutional reports (see Appendix 2i). In accordance with an earlier plan, these institutional reports were distributed to CAEL Assembly members for experimental use along with the working papers.

The Validation Plan

These working papers and institutional reports produced during 1974-75 included descriptions of competences, suggested assessment procedures, specimen materials, etc. In the following year this work moved into a second important phase--the tryout and validation of these materials in CAEL Assembly institutions. While the CAEL Steering Committee regarded these materials as an important step forward and quite suitable for experimental use, the materials clearly did require review and could, undoubtedly, be improved in a variety of ways. Part of that improvement would be description of additional applications at various institutions of the general assessment methods and principles initially described. So an important outcome of the validation phase was to be the revision, improvement, and actual production of formal published materials based upon the earlier experimental reports. The CAEL Assembly includes a broad cross-section of American higher education, and the plan for directly involving these institutions in the validation of developmental work represents a unique model for insuring that CAEL outcomes are sound and for facilitating direct benefit of CAEL work in many institutions.

Another part of the validation work is more technical in character and refers to the process of verifying that the assessment materials and procedures are adequate from an educational and measurement standpoint--especially that they have satisfactory validity and reliability. This raises an important question, both practically and theoretically, as to what constitutes establishing validity and reliability with respect to this type of assessment.

An obvious problem lies in the fact that CAEL did not produce and had no intention of producing a set of standard instruments that could be readily subjected to statistical criteria of technical adequacy. As already noted, CAEL staff and institutions have produced principles, guidelines, suggested procedures, prototype documents, and methods that are generally applicable to sound assessment in a variety of learning situations. Establishing the validity of a standardized test is difficult in any case; procedures for assessing experiential learning are overwhelmingly more complicated, as must be their evaluation. There are, however, important ways in which the educational soundness and the technical adequacy of CAEL materials and procedures can be put to the test and selectively supported or reinforced as appropriate. The following ways are especially important:

I. Review

- Through a systematic process of critical review by all parties likely to be concerned with or affected by assessment (students, faculty, outside experts) in order to determine that the procedure and content of assessment are appropriate.

II. Experimental Use

- Through careful evaluation by practitioners and experts as to whether quality control procedures are adequate and whether practices proposed reflect proper priority in order to produce sound and equitable assessment.
- Through systematic development of local "adaptations" in order to fit assessment procedures to the character and demands of local situations.

III. Field Research

- Through empirical studies that help to determine the educational and technical quality of different assessment methods and to reveal important characteristics of assessment procedures that tend to enhance or detract from reliability, validity, or usefulness of local applications.
- Through the development of validation models whereby individual institutions can monitor quality assurance of assessment procedures, especially with respect to the consistency of assessment and to achieving consensus regarding its content and process.

These several approaches reflect the need to establish the appropriateness of the assessment principles suggested in CAEL materials and also to show how local programs can insure that their own applications are sound from a technical standpoint. Since the CAEL Assembly includes a diversity of colleges, universities, and agencies, different Assembly members use CAEL experimental materials in different ways. Accordingly, three different levels of participation were arranged for institutions wishing to engage in validation activity. The three levels of participation in this work correspond to the three types of validation activity outlined above. Some Assembly members were engaged in critical review (Level I) and used CAEL materials primarily as a basis for staff discussion and planning. Other member institutions used student and faculty guides experimentally (Level II). Still other institutions had a more substantial involvement with the project as field research sites (Level III). These three levels were sequentially inclusive; that is, CAEL member institutions were offered three levels of participation defined as follows:

Level I	Level II	Level III
Review	Review	Review
	Experimental Use	Experimental Use
		Field Research

These three levels of institutional participation in CAEL validation work were first described in a "Prospectus" mailed to all CAEL member institutions in the spring of 1975 (see Appendix 2a). In early June a formal invitation was sent to CAEL member institutions (see Appendix 2b). All were invited to participate at Level I or II and encouraged to apply for Level III where participation was limited. An external committee (Leland Medsker--Chairman, William O'Connell, and Patricia Thrash) selected the institutions to participate at Level III (see Appendix 1c).

As previously suggested, these different types of institutional participation formed the basis of two forms of validation. One was review and experimental use of materials that contributed to a consensual validation and revision of assessment reports. The nature and outcome of that work is described in the following chapter. The field research produced the second form of validation--extensive empirical studies of assessment principles and practices. The nature and outcome of that work is described in Chapters III through VII. The following section provides a very brief guide to what can be found in the following chapters.

Overview of This Report

Chapter II describes how the CAEL reports were used experimentally and reviewed, the feedback the reviews produced, and what important issues were raised in the process. It also summarizes the types of changes that were made in the assessment reports as a result of the year-long process of review, experimental use, and feedback. This chapter stands alone and can be skipped over if the reader's interest lies in the results of the field research.

Chapter III starts with some discussion concerning the reasons for carrying out certain types of validation studies and provides a general description of how the field research was carried out. Thus, this chapter provides background for the three that follow. Readers who are unfamiliar with correlational statistics and how they apply to questions concerning reliability and validity are especially encouraged to read the final section of Chapter III on "Interpreting the Results" before getting into the actual findings of the field research. That same section also provides some useful information concerning important qualifications that need to be borne in mind.

Chapters IV, V, and VI report on the procedure and the findings of the ten studies that were undertaken in Areas A, B, and C, respectively. Each of these three chapters is self-contained, so readers can feel quite free to read any or all as their particular interest may dictate.

Chapter IV is concerned with the assessment of interpersonal skills and should be of special interest to readers interested in the special problems of assessing competences that involve complex performance not clearly defined in a traditional academic frame of reference. Chapter V reports several studies concerning portfolio assessment and is especially relevant to readers who are interested in assumptions and practices involving the crediting of prior learning. Chapter VI reports on three studies involving the use of the "work assessment model" to award credit for highly specific work-related competences that correspond to requirements in occupational programs.

Chapter VII is quite different from the previous three. It includes several exemplary institutional studies of the nature of faculty agreement in the assessment process. This work was not intended to test principles or to produce generalizable findings regarding assessment practices but rather to illustrate the sorts of studies that can contribute importantly to institutional quality assurance. These studies in Chapter VII are recommended as useful models.

Chapter VIII provides a concluding overview of the validation phase of CAEL. It attempts to identify assessment strengths and weaknesses and comments on those areas where assumptions and practices appear reasonably sound or questionable. This final chapter is especially oriented toward implications--that is, what this validation work suggests with respect to developing programs and improving practices, assuring quality and equity, and fostering assessment methods that benefit students as well as institutions.

Chapter II

REVIEW AND REVISION OF CAEL MATERIALS

Jane Porter Stutz and Warren W. Willingham

As explained in Chapter I, CAEL validation work included two general types of activities: review and revision of materials developed during 1974-75 and a series of field research studies. This chapter provides an overview of the former. Materials developed in 1974-75 included four assessment handbooks (a fifth was subsequently added), three student guides, and five institutional reports. The content of these various reports is outlined in the following section. They were prepared by individuals from eleven CAEL institutions and ETS staff (see Appendices 2i and 2j). Subsequent sections of this chapter give an overview of the process in which these reports were tried out and revised during 1975-76 and a brief account of the outcome of that process.

Description of the Reports

Area A - Assessment of Interpersonal Skills

The working papers in this area--one for faculty and one for students--were based upon an analysis of the interpersonal domain originally by Sidney Fine. He and his associates developed a classification system for occupational roles, differentiating them according to their demands for work with people, with data, or with things. Within this classification Fine suggested a number of subcategories; those relating to people were taken as a basis for a CAEL adaptation to experiential learning with a focus on the important societal roles which demand interpersonal competence.

The faculty guide, Working Paper No. 4, provides a general framework for understanding interpersonal skills and their relation to undergraduate programs. The student guide, Working Paper No. 5, translates these ideas to the student's needs and point of view and provides illustrative cases from institutions offering experientially-based credit for interpersonal competences.

In addition to a theoretical structure, these two working papers provide a description of strategies for developing a program and for assessment, the latter based in part on prior work presented in CAEL's Compendium of Assessment Techniques. The working papers explore opportunities for applying these assessment methods to learning in the interpersonal domain. Furthermore, a general purpose model is suggested that includes alternate procedures for adaptation to particular assessment situations on individual campuses.

Area B - Use of the Portfolio to Assess Non-Sponsored Learning

This area included two working papers: a faculty guide and a student guide. The faculty guide, Working Paper No. 6, set out to answer three questions that faculty might well raise before embarking on the assessment of prior (or non-sponsored) learning: 1) what are the types of prior experiences and learning outcomes resulting from these experiences that are presented to faculty for assessment? 2) What portfiolio processes are now used to assess prior learning? 3) What choices do faculty have concerning general steps and alternate procedures that might be used in portfolio assessment?

The guide addresses the first two questions by providing information about the state of the art in portfolio assessment. Approximately 400 portfolios were examined during more than 30 institutional site visits. The third question is addressed by describing various alternate procedures in each stage of an eight-step model of portfolio assessment. Sample checklist forms, outlines, documents, and instruments from cooperating institutions are included to illustrate various procedures. Advantages and disadvantages of alternate procedures are also discussed.

The student handbook, Working Paper No. 7, is written in an informal style and is designed for use by adult students entering or returning to higher education. It takes into account a variety of institutional assessment procedures and arrangements. The situations range from institutions that have no special program for prior learning and assessment to institutions that have credit courses in educational planning and learning assessment. It describes clearly and simply the steps in portfolio assessment and uses illustrations of six students who are seeking university degrees. Prior learning experience is described for each of the students, and the handbook user is set to the task of applying the portfolio assessment procedures to the examples and eventually to him- or herself. This document is designed not only to help the student with these practical problems concerning the construction of a portfolio but also to guide the student through what can be a highly educational experience.

Area C - Assessing the Learning Outcomes of Work Experience

This area includes two independent pieces of work: Working Paper No. 8, A Task-Based Model for Assessing Work Experience, and Working Paper No. 9, A Student Guide to Learning Through College-Sponsored Work Experience. Subsequent to the publication of the latter report, a companion faculty handbook was prepared.

Working Paper No. 8 presents a general model for assessing specific competences in occupational fields, especially those required in career related degree programs. The model is designed to help a college specify the kinds of skills and competences acquired in various occupational settings, to define the learning outcomes of its occupational and career programs, and to translate students' relevant work experience into college credit. In seven sequential stages the model describes how lists of competences can be developed in an occupational field and how they can be used for various objectives in higher education that require the assessment of work experience.

Working Paper No. 8 includes a detailed application of the Work Assessment Model to the fields of data processing, law enforcement, and secretarial science. In each of these fields a number of prototype assessment instruments or procedures have been developed on the basis of compendia of competences. Thus, in these three fields there are products that could be immediately applied in a practical situation.

Working Paper No. 9 is a guide for students who are participating or are planning to participate in sponsored experiential learning programs (work-study programs, internships, cooperative programs, etc.). It was designed as a combined planning guide, study outline, and self-evaluation device, to help students meaningfully integrate their work experience into an overall educational program. The student guide discusses such issues as career exploration, acceptance of responsibility, preparation for the learning experience, involvement in the experience, and synthesis of the learning following the experience. It recognizes differences between various types of off-campus programs, discusses the use of self-ratings and other ratings, and directs students to appropriate information sources on their own campuses.

More specifically, the guide outlines a series of 11 steps through which the student is encouraged to progress. These range from taking stock of goals and considering available work options to assessing learning outcomes after the off-campus experience and identifying new learning objectives. These steps are accompanied by appropriate worksheets and checklists. Following the preparation of this report, it was recognized that such a student guide would be greatly strengthened by a handbook for faculty, staff, and field supervisors. Such a handbook was prepared and is to be published as part of the revised series of CAEL assessment reports (see Duley/Gordon report, Appendix 2j).

Area D - The Use of Expert Judgment in Assessing Learning Outcomes

Due to the highly individual character of much experiential learning, assessment often must rely upon the informed judgment of an expert. Working Paper No. 10 provides principles and guidelines for faculty, students, and consultants on the use of expert judgment. The first portion of this report deals with procedural issues that cut across the various contexts in which expert judgment might be applied; for example, 1) functions of the expert assessor in addition to actually judging competence (defining criteria and structuring and adapting the assessment procedure to the student), 2) the importance of establishing assessment standards that are defined as systematically as possible and related to objective, observable indices or behaviors, 3) the issues of reliability and validity, illustrated with examples of rating errors that bear on the credibility and the fairness of the assessment, 4) evaluation of assessment, and 5) general principles and steps in sound assessment.

These various issues, principles, and steps in assessment are illustrated and discussed in a series of chapters covering the use of judgment in four areas: interviews, product assessment, performance assessment, and assessment of written material. Discussion of each of these four areas in which expert judgment is commonly applied includes a number of practical hints and suggestions for the improvement of assessment as well as problems and pitfalls to avoid. Several applications are described in some detail; e.g., prototype assessment interviews, a model for product assessment in the visual arts, an institutional plan for performance assessment, and an assessment model for off-campus experiential learning programs.

Institutional Reports

While the seven working papers represent the major outcome of the developmental work at the core of the project, another set of papers resulted from special projects carried out on individual campuses. Twenty such projects, in the same four priority areas as the working papers, were funded by CAEL in 1974, and five of the project reports were selected for publication as CAEL Institutional Reports. These reports were distributed in the fall of 1975 and were included, along with the working papers, in the process of critical review and experimental use.

Institutional Report No. 1, The Use of Expert Judgment in the Assessment of Demonstrated Learning in the Antioch College Yellow Springs Adult Degree Completion Program, describes the program and summarizes workshops designed to help inexperienced faculty understand the special problems involved in evaluating demonstrated learning of adults. The workshops provided opportunity for faculty already involved in the program to share their experiences, discuss ways of improving their evaluation techniques, and facilitate the evaluative activities of their colleagues. Case studies, articles about specific practices on the Yellow Springs campus, sample demonstrated learning evaluation reports, and samples of actual forms are included.

Institutional Report No. 2, Interpersonal Learning in an Academic Setting, was prepared at Empire State College with major emphasis on the incorporation of interpersonal learning into the formal educative process. At Empire State learning is structured via a learning contract developed by the student with the help of a mentor. This paper demonstrates how the learning contract mode can be and has been applied to interpersonal learning through designing learning activities and assessment processes. Findings are interpreted in the context of a learning framework that can take account of cognition, affect, and behavior (thought, feeling, and action).

Institutional Report No. 3, Evaluation and Expert Judgment, was prepared at the University of California at Los Angeles and centers around the development and use of a record-reflection log and a process-assessment matrix as tools to be used in assessing experiential learning. The text of the report describes the stages of the project, and the appendix includes all the relevant documents, including a tentative "Guide to Assessing Experiential Learning." Because faculty have much less direct contact with and supervision of students in experiential learning situations than in traditional classroom settings, the report explores possible roles for other persons--especially the student and an independent evaluator--in the evaluation process. It also stresses formative evaluation for monitoring the progress of both the individual student and the course or program as a whole.

Institutional Report No. 4, The Refinement and Modification of an Instrument for Assessing the Achievement of Interpersonal Skills of Social Work Students, was prepared at Wayne State University. It describes the development of a Performance Assessment Checklist and its use in a field work practicum in social work. The project and checklist grew out of a recognized need for a behaviorally specific instrument to assess student performance in the field based on the identification of specific behaviors and activities of students that reflect their learning. It is intended to counteract some undesirable assessment practices, such as ambiguous reports, or assessments presented largely as personality dynamics.

Institutional Report No. 5, Guidelines and Procedures for the Assessment of Experiential Learning and for the Selection and Training of Field Experts, describes how William Rainey Harper College established tentative procedures for assessing the non-sponsored experiential learning of students applying to a new, nontraditional Associate in Liberal Studies Program. A task force which represented a cross-section of the institution met with consultants, conducted site visits to other institutions that had programs for assessing experiential learning, and eventually produced two documents: a set of procedures and a student guide for portfolio development. In addition, the task force created a seminar designed to help students develop their portfolios.

The Tryout Process

The purpose of the tryout process was to produce revised and improved assessment reports, but the strategy of the tryout was more complicated. As previously noted there was considerable concern both in and outside the project that CAEL not produce premature solutions to assessment problems. Consequently CAEL encouraged all CAEL Assembly institutions to participate in the tryout process. In order to facilitate such participation, it was formalized with invitations, offers of free CAEL reports, and an agreement that feedback would be provided in return.

Two outcomes were desired from the extensive participation: 1) CAEL sought useful reactions and advice concerning the content of each report, i.e., weaknesses, important issues, needed development, and 2) CAEL sought to stimulate as many institutions as possible to become engaged in the assessment problem by extending the present work, creating alternatives, and adapting particular aspects of CAEL procedures to the demands of local situations. Thus the general plan evolved: engaging the institutions, inviting them to submit their own work, obtaining critical feedback and suggestions (much came informally to the authors), identifying issues and suggested revisions, discussing those issues at the spring 1976 CAEL Assembly meeting, incorporating new material in revised reports as seemed appropriate, and producing the revised reports (see schedule, Appendix 2c).

In June, 1975, an invitation to participate in CAEL activities and tryout of materials (working papers) was extended by the project director to all representatives of CAEL 1975-76 member institutions (Appendix 2b). Interested institutions submitted an application for participation in the tryout and indicated whether they wanted to be involved at Level 1 (review and critique) or Level 2 (experimental use), which of the four areas they were interested in, and the number of participants (faculty and students) they expected. On the basis of these forms, institutions were sent an appropriate number of gratis copies of the working papers. The agreement was that all participants would return a brief questionnaire in the spring of 1976 to report the results of their critical review. Level 2 institutions were asked to use the materials experimentally in programs already under way and, if they adapted the materials, they had the option of submitting the adaptation as an institutional model for possible use as an illustration of an exemplary procedure. Applications were received throughout the summer and fall.

In October a similar invitation was extended to review and try out the five institutional reports. Some 80 institutions signed up to participate in the tryout of working papers and institutional reports. Since participation was sequentially inclusive, Level 3 field research sites were also participants at Levels 1 and 2. A list of participating institutions is included in Appendix 1d.

A basic evaluation form, designed to be completed by faculty or administrators, was developed which would apply to all working papers and institutional reports. In addition, a student form was developed for the reports where student feedback would prove useful. The authors of reports were given an opportunity to add additional questions unique to their individual reports. Because of the addition of the unique questions, 16 different two-page forms were developed (see Appendix 2g)--one faculty form for each of the 12 reports and four student forms. Multiple copies were distributed in February 1976 to participants and a large number of other individuals working on CAEL activities. Returns were requested by March 15. CAEL did not specify the manner in which the institutions should conduct their reviews, critiques, and experimental tryout. Most feedback came in the form of completed evaluation forms; some respondents also included formal reports of the tryout process on their campuses; some sent only brief replies. Many people also communicated directly with the authors. It was through such contacts that most illustrative material from other institutions was added to the revised reports, though the project director received a number of adaptations and special reports in response to the invitation for such materials. The overall impression is that the reports were widely read and used and that they stimulated much discussion with the authors and among other members of the CAEL Assembly.

Outcome of the Tryout

The outcome of the review and experimental use of CAEL assessment materials can be described under three headings: 1) the evaluative feedback provided formally through the questionnaires sent to participating institutions and informally through individual contacts between authors and interested persons; 2) the changes that were made and the material that was added to the original assessment reports largely on the basis of formal and informal evaluative feedback; and 3) adaptations, extensions, and other work stimulated by these materials and by CAEL work in general. These different types of outcomes are summarized in the following paragraphs.

Evaluative Feedback

While the authors of the original reports devoted a good deal of time during the year to discussing their work with colleagues and examining materials and ideas for possible inclusion, we can only report here on a more limited aspect of the evaluative feedback--the formal feedback returned to CAEL on the evaluation forms. Some 400 were received; Table II-1 gives an indication of their source.

When the evaluation forms were received at ETS, copies were sent to the authors of the pertinent report so that they could make note of specific criticisms and suggestions and also identify from all feedback received those important issues that needed further discussion prior to final revision of the reports. The authors identified what they considered to be the most important issues (see Appendix 3a), and those issues served as the agenda for special sessions at the spring 1976 meeting of the CAEL Assembly in Denver.

Table II-1: Number of Forms Received and Percentage of Respondents Who Judged the Reports Useful*

	REPORT							
	A	B	C	D	E	F	G	Total
FACULTY RESPONDENTS								
Type of Institution								
2-year	7	2	6	11	7	13	6	52
4-year	24	38	35	32	28	24	33	214
Public	15	17	21	30	20	23	21	147
Private	11	20	17	9	13	9	17	96
Special Interest in Assessment								
Experiential learning comparable to course work	8	13	16	25	9	19	10	100
Off-campus learning and internships	20	17	17	14	18	13	24	123
Curriculum innovation and different learning outcomes	16	19	19	18	20	15	20	127
Institutional Interest in Experiential Learning								
Recent	9	12	--	20	8	18	8	75
Several years old	22	27	--	22	27	18	29	145
All Faculty Respondents								
Number of forms	30	41	42	43	35	38	38	267
% who judged report useful*	70%	61%	76%	86%	63%	74%	82%	73%
STUDENT RESPONDENTS								
Number of forms	20	--	--	--	48	26	--	94
% who judged report useful*	65%	--	--	--	73%	88%	--	76%
TOTAL: Number of forms received	58	45	44	46	88	75	45	401

*"Useful" includes all responses in the categories "very useful" and "useful or somewhat useful."

A content analysis of common questions was undertaken by CAEL staff. The following two questions were common across all feedback forms:

1. What would you identify as particular strengths or weaknesses in this report?
2. Everything considered, do you feel that this report is useful or not useful in your particular institution (or educational unit)? Do you have suggestions for improvement?

A third question appeared on the faculty forms for all reports:

3. Can you cite specific issues concerning assessment practice or policy that bear on the revision of this report and might be profitably discussed at the spring CAEL Assembly meeting?

The purpose of the content analysis of these questions was three-fold: 1) to gain perspective on how the whole review process worked out and its implications for the reports in general, 2) to provide individual authors with a useful frame of reference for interpreting reaction to their reports in comparison to respondent reaction to other reports, and 3) to aid the Executive Committee in review of manuscripts in late June. The three general questions and those other questions added to individual feedback forms by the authors naturally elicited many suggestions and comments specific to unique aspects of individual reports, but the content analysis was not intended to summarize such comments.

The content analysis involved these steps: For each question two readers established categories of responses that were relevant across reports. Then all responses to each question were coded according to the established categories and tallied on five tables. Table II-1 shows for each working paper the percentage of respondents who found the reports useful. Tables in Appendix 3b show the percentage of respondents who cited different kinds of strengths, weaknesses, suggested improvements, and critical issues. Because too few forms were received on institutional reports, they were not included in the analysis.

In looking at the tables the reader should be aware that the percentages reported are frequently based upon samples of 40 or fewer respondents. This means that the percentages are not very stable, and small differences among percentages should be ignored. On the other hand, the questionnaires were open-ended, so there may be significance in even a limited number of people volunteering the same type of comment.

In general, the feedback was quite favorable. About three-quarters of the respondents found the materials to be useful to some extent. In cases where the materials were judged not useful, it was usually for one of two reasons. First, some reports were relevant to certain types of programs and were not intended to be useful to all institutions. Second, some respondents did not understand that the materials were designed to be adapted to local needs; i.e., it was expected that institutions would select portions of the working papers that were relevant to them and perhaps add materials of their own to be used in conjunction with the CAEL reports.

The strengths and weaknesses question most often elicited specific and useful comments regarding particular sections of reports. For example, students offered such comments as the following about the student guide on interpersonal skills: "The Appendix is very helpful." "From a student point of view, it seems the most useful chapter would be Strategies for Interpersonal Skill Development with its questions, exercises, and activities tied together in the Planning section." "I would recommend some reworking of the chapter on the Analytic Framework in terms of its usefulness to students."

Students and faculty made a number of such useful comments and suggestions, and most of the authors found the evaluation forms very helpful. The content analysis itself, however, proved to have limited usefulness. It did show the diversity of responses and the contradictory nature of many of them. It was common to find one evaluation form saying that a given paper was well organized and very readable and another form on the same paper characterizing it as poorly organized and difficult to read! If the analysis showed anything, it documented the impossibility of pleasing all segments of a diverse audience. The more important finding, however, was the fact that quite substantial numbers of readers found the reports useful for their purposes.

Significant Revisions

Three types of CAEL reports were tried out experimentally during 1975-76: four faculty guides, three student guides, and five institutional reports. In each case, the author or authors of the original report produced a revised manuscript for possible CAEL publication though the extent and nature of the revision differed considerably from report to report. The general nature of these changes for each of the three types of reports can be summarized as follows.

The series of four working papers that originally had been called "Faculty Guides" were termed "Handbooks" in revised form in recognition of the fact that those reports were appropriate for other institutional staff, field supervisors, etc. in addition to faculty. The most significant change in this series of reports was the addition of a fifth handbook due to a gap in the series that was considered significant. Thus, the handbook on sponsored learning (by John Duley and Sheila Gordon) was the result of a special added effort during the year.

In the handbooks that were revised, major changes tended to fall into three categories: revisions in the original substance of the report, addition of new material, and improvements in the usefulness of the report to the reader. The revised handbook on interpersonal skills provides two examples of substantive revisions: a more fully developed description of interpersonal literacy and additional emphasis on the assessment of interpersonal competence in sponsored programs. A substantive revision in the portfolio assessment handbook was the increased attention given to the rationale for crediting prior learning and the context in which the assessment occurs. In the handbook on expert judgment, an example of revision in original substance is the chapter on product assessment, which was rewritten to be more practical.

In addition to substantive revisions of old material, these reports were extended by addition of new material. For example, the report on interpersonal competence includes new material connecting that work to major lines of theoretical development in this area and a new chapter specifically on the interview as a means of assessing interpersonal competence. The revised report on prior learning gives special attention to the role of faculty. Furthermore, all the handbooks now stress the importance of evaluating assessment and the implications of CAEL field research for quality assurance.

The handbook authors attempted to use advice from the evaulation questionnaires to improve the utility of reports wherever possible. Unnecessary material was removed, more useful materials and illustrative cases were added to appendices, and detailed examples were added to show how assessment models could be used (e.g., a portfolio model and the work assessment model). Authors also gave attention to the problem of helping faculty use the student guides effectively in conjunction with the handbooks.

In the revision of the student guides there were three main types of changes: substantive revisions or extensions, emphasis on the "how-to" quality of the guides, and changes in design and format to make them more useful and interesting to students. Substantive revisions included a new typology to help students understand the various types of field experience, a more practical orientation in the description of interpersonal competence, and the addition of illustrative materials to help students understand better the portfolio assessment process.

There was a special effort to make the student guides more directly useful by including more summaries of principal points, more checklists, more examples, and more worksheets. Wherever possible, authors attempted to express their main points as steps with alternatives as appropriate.

Authors acted on a number of suggestions for improving the design and format of the student guides. Students were quick to point out jargon and extraneous material and a somewhat dull appearance of the original versions. The editors tried to improve the format and layout of the guides so that they were more readable. Finally, in designing the format special account was taken of the fact that some students may wish to use the material in a notebook that includes additional local forms, instruction, materials, etc.

Each of the five CAEL institutional reports published during 1975 was revised or extended on the basis of work during the year. After careful review, however, the CAEL Executive Committee judged only two to incorporate sufficient new material to warrant new publications. These were the reports prepared at Empire State College and the University of California at Los Angeles. It was decided that the other three reports would be retained in print as long as demand indicated.

Adaptations and Extensions

One of the purposes of the validation phase of CAEL was to try out the experimental materials in a number of quite different situations--partly to test the appropriateness of suggested procedures but also to encourage modification and adaptation of general assessment procedures and prototype materials to the special conditions of different learning situations and populations of students. CAEL actively encouraged the development of local models in order to enhance the appropriateness of assessment procedures and to avoid undesirable standardization of techniques and instruments.

Institutions were specifically invited to submit "adaptations" of the CAEL materials to be considered for possible inclusion in the revised versions of the reports. At that time an adaptation was described as a brief paper that (1) focused on a specific aspect of assessment related to one of the 12 CAEL reports, (2) described a useful variation from procedures outlined in the CAEL report, and (3) gave a rationale for the variation in terms of different local circumstances, different program objectives, improved efficiency or effectiveness, etc.

The intention was to increase institutional involvement in CAEL development and broaden the assessment reports by adding material that would be relevant to a range of situations. In several important respects these objectives were realized. Certainly CAEL helped to stimulate a number of institutions to develop their own materials and to break new ground on new assessment problems. The large number of proposals for contributed papers submitted to the CAEL Assembly Program Committee for the spring 1976 meeting is witness to that fact. Furthermore, the revised reports do include a variety of useful illustrative materials, forms, examples, and so on, from various institutions that CAEL authors located and included with appropriate acknowledgment of the sources.

Since CAEL had encouraged "adaptation" of these preliminary materials, there were several instances in which institutions used the CAEL reports with minor modifications so as to make them more appropriate for use with local students. Whereas CAEL had no objection to use of such locally modified versions during the experimental year, these were not the "adaptations" hoped for. In some cases the materials were not adapted because the programs in which they were used were so new that most time and efforts were still being devoted to program design and implementation. These institutions welcomed the CAEL materials as guidelines and procedures that could be used immediately as a place to start. They simply were not yet in a position to provide the kind of adaptation we had hoped for. It is also possible that for most participating institutions the definition of "adaptation" that CAEL had put forward simply turned out to be too abstract and did not readily apply to the many ways in which the materials were being used on campuses. Nonetheless the very considerable amount of developmental activity in scores of institutions across the country provides reassurance that basic assessment principles are being adapted to local circumstances and there is little if any danger of premature solutions or standardized methods in the assessment of experiential learning.

Chapter III

OVERVIEW OF CAEL FIELD RESEARCH

Warren W. Willingham and Kurt F. Geisinger

CAEL field research was keyed to the four developmental areas previously described. In areas A, B, and C field research consisted of a series of ten multi-institution validation studies based upon data collected by the institutions according to predetermined specifications and subsequently analyzed and reported by ETS staff. This chapter and the three that follow are concerned with these ten studies. Research in Area D was a different type--a series of local studies of consistency of judgment. Three of these institutional studies are reported in Chapter VII.

While the ten validation studies reported in the next three chapters focus upon different types of learning or assessment, they have a great deal in common with respect to purpose and procedure. Consequently, this chapter provides necessary background information that is helpful in understanding the nature of the studies, why they were carried out, and important points to remember in interpreting the results.

Since this report is intended to be a document of record, it contains a great deal of detailed information, some of it specialized. The intended audience is diverse, some being interested in analytic procedures and others not. In order to make the report complete but also comprehensible and manageable, most of the technical and procedural detail has been put in the appendix. The body of this report assumes no special sophistication in statistics or research design though some such knowledge will undoubtedly help in interpreting some of the results which are necessarily technical in nature. The last section of this chapter attempts to orient the reader to the main types of analysis reported here.

Purpose and Objectives

Chapter I provided a general rationale for the validation phase of CAEL and differentiated two types of validation activities: a) critical review and experimental use of materials and b) validation studies carried out through field research. The purpose of the former is clear enough and the outcome of that process has been described in Chapter II. The purpose of the validation studies is more complex and requires special consideration. The following paragraphs engage these questions:

-- Why are validation studies needed?

-- What are the questions that validation studies can help answer?

-- What specifically is being validated?

-- How are the studies being carried out and what strengths and weaknesses of the field research plan need to be taken into account?

Validation studies are important for two principal reasons. First, systematic research is needed as a basis for suggesting revisions in assessment methods already suggested by CAEL. More generally, this research will contribute to the continuing development of benchmarks for improving assessment practices. Second, in addition to the obvious need for assessment methods to warrant the confidence of educators who use them (especially with respect to substance and feasibility) it is also necessary to meet the usual requirements of acceptable measurement practice if credit earned through experiential learning is not to represent discounted currency in higher education.

That is, if credit for experiential learning is to be broadly acceptable, it will be necessary to develop especially firm quality control regarding the assessment process. This means demonstration of adequate reliability and validity of evaluation methods.

It is important to recognize that reliability and validity are not statistical abstractions but represent the most fundamental basis upon which one decides whether assessment is defensible. For "reliability" one might well read "adequate consistency in assessment to insure equity to individual students." Validity frequently means reasonable agreement among interested parties concerning what learning outcomes are required of the student. Furthermore, it is highly desirable to demonstrate that good assessment is in itself a beneficial learning experience for the student.

Favorable testimony is not sufficient for these demonstrations. They are complex questions that require careful study and good evidence. The principal purpose of the field studies is to provide the concrete data that are helpful and necessary in showing whether the assessment process is consistent and fair and in what respects it needs improving in the interest of students and institutions. The following brief outline illustrates the sorts of specific questions to which the validation studies will provide relevant data and some partial answers. Each question is keyed (A1, C3, etc.) to studies described in the following chapters.

I. Validity of assessment

Do faculty agree reasonably well on:	
what is creditable?	B2
how competencies are related to degree programs?	A1, C2
what documentation of learning is required?	B3
Is documentation of learning typically authentic?	B3
Is learning reported by students verified by work supervisors?	C3

II. Reliability of assessment

Are faculty reasonably consistent:
  in judging the level of competence students have reached? A2, B1
  in using alternate methods of assessment? A2
  in awarding credit for the same portfolio or petition? B1, C1

III. Educational value of assessment

Is there a discernible benefit of assessment to the student? (i.e., is assessment a useful learning experience?) A3, B4

Another way of looking at validation is to ask what is being validated--instruments, reports, procedures? It is clear that CAEL is not producing any instruments in the sense of standardized tests, but prototype forms included in the assessment materials can be thought of as instruments. On the other hand, none of these instruments or working papers are necessarily fully applicable to a particular institution or program. It is more accurate to say that the field research is concerned with the validation of procedures and that the material in the working papers is very much part of that process.

Through the developmental work reported in the working papers CAEL is fostering certain assessment procedures on an experimental and tentative basis. These procedures incorporate a variety of principles and guidelines that are suggested by a) experience and careful analysis of the many individuals involved in the developmental work and b) technical understanding of assessment built up through the long history of educational measurement. It is the reasonableness and feasibility of these procedures that is being put to the test in the validation studies.

Finally, it is useful to differentiate two types of studies. Some studies test assumptions that are implied by the assessment procedure outlined in the working papers (e.g., the assumption that faculty members can agree reasonably well as to what types of learning are creditable). That type of study is often based upon illustrative materials. The second type of study tests assessment practices (e.g., whether faculty are reasonably consistent when they award credit for learning described in a particular portfolio). Appendix 2e outlines briefly the use of working papers in relation to the studies.

Description of the Studies

The preceding discussion was concerned with the general rationale of the validation studies and a description of how they fit into the overall CAEL project. Outlined below are brief descriptions of the 10 individual studies in Areas A, B, and C. It will be noted that each of these studies refers to one or more of the stages of assessment to which CAEL materials frequently refer:

Stage 1--Identify the learning acquired or intended
Stage 2--Articulate learning to educational goals or academic degree
Stage 3--Document learning experience or outcomes
Stage 4--Measure the extent and character of learning
Stage 5--Evaluate whether learning meets an acceptable standard
Stage 6--Transcribe the credit or recognition

The purpose and character of these individual studies is indicated in three ways: by the marginal notation of the stage of assessment and the type of technical problem primarily involved, by the brief narrative explanation of the rationale, and by an indication of the data each institution was required to collect.

AREA A - Assessment of Interpersonal Skills

Stage 2
(Validity)

Study A1: To what extent can faculty articulate interpersonal competences with different degree programs?

If there is a clear rationale for crediting interpersonal skills, then there should be acceptable agreement among the faculty with respect to which interpersonal competences are relevant for various degree programs. The basic data were judgments of faculty at each field research site as to which interpersonal competences were relevant to each of several selected degree programs. The interpersonal competences to be judged were presented in the form of 15 illustrative cases.

Stage 4 (Reliability)

Study A2: To what extent do different assessors agree on the level of interpersonal competence students have acquired?

In order for assessment of interpersonal skills to be fair and valid there should be reasonable agreement between different assessors using the same method and between assessments based upon different methods. Assessments of the level of interpersonal competence were made for samples of students on each campus using a structured interview and a role-playing technique. For each of these methods, two judges made independent assessments.

All Stages (Educative Value of Assessment)

Study A3: What is the educational value to the student of the assessment of interpersonal skills?

If assessment of interpersonal skills has educative value, then a significant difference should be observed among students who have and have not gone through the assessment; viz, students who have experienced the assessment process should have a better sense of their own interpersonal competences and the relation of those competences to their personal goals. Free-response answers to questions concerning student understanding of their competences and their educational goals were obtained for a sample of students prior to assessment and for another sample of students after assessment had taken place. Four faculty judges rated these responses (blind) for quality of manifest self-awareness.

AREA B - Portfolio Assessment

Stage 5 (Reliability)

Study B1: Do faculty agree on the amount of credit awarded for portfolios?

In order to be fair and equitable, crediting practices must be consistent. One way of evaluating the consistency of crediting practices is to determine how closely assessors agree on the number of credits to be awarded after independently examining the same evidence. Two different types of data were collected. On each campus a set of local student portfolios were assessed twice by the method used locally. A second type of data consisted of two independent faculty judgments of credit that might ordinarily be awarded to each of 14 illustrative portfolios adapted from actual portfolios but altered so as to ensure anonymity.

Stage 1 (Validity)

Study B2: Do faculty agree on what constitutes college-level learning?

Students submit various types of experiential learning in petitions for college-level credit. A basic requirement for the validity of crediting practices is that faculty agree reasonably well on which types of learning are creditable. In this study each institution provided judgments by 5 faculty as to whether each of a series of 15 illustrative examples of learning was considered creditable as college-level learning.

Stage 3 (Validity)

Study B3: To what extent do faculty agree on the type of documentation required as evidence of experiential learning, and to what extent do various types of documentation prove authentic?

Students often seek credit for experiential learning which should and can be substantiated with documentation. This study examined the degree of consistency among faculty with respect to the amount and nature of documentation they judge to be required for various examples of experiential learning. Also, because of limited resources, it is often not possible for thorough checks to be made on the authenticity of all documentation. It is possible to make more thorough checks for a limited number of cases for research purposes, however, and thereby estimate the authenticity of various types of documentation. A set of 15 illustrative cases were presented to five faculty who were asked to indicate in each case what types of documentation were required. Also, documentation of various types in 25 actual student portfolios were checked for authenticity.

All Stages
(Educative
Value of
Assessment)

Study B4: What is the educational value to the student in developing and having a portfolio assessed?

If portfolio development and assessment have educative value, a significant difference should be observed between students who have not gone through this process, especially with respect to self-awareness of personal competences and educational goals. Free-response answers to questions concerning student understanding of personal competences and educational goals were obtained for samples of students prior to assessment and other samples of students after assessment had taken place. Such material was judged blind by four faculty judges according to pre-determined criteria of self-awareness manifested.

AREA C - Assessment of Work Experience

Stage 5
(Reliability)

Study C1: Do faculty agree on the amount of credit to be awarded for work experience?

If credit offered for work experience is to be fair and equitable, crediting practices must be reasonably consistent. One way of evaluating the consistency of crediting practices is to see how closely assessors agree on the number of credits to be awarded after independently examining the same evidence. In this study credit for occupational competences was evaluated with a local adaptation of the Work Assessment Model (Working Paper No. 8). Work experiences of students in several institutions and fields were assessed independently by 2 local faculty judges.

Stage 2
(Validity)

Study C2: To what extent can faculty match work experiences with relevant requirements in an occupational program?

If practices for crediting prior work experience are sufficiently well articulated with the curriculum, faculty should be able to assign specific competences to relevant courses in a particular degree program with reasonable consistency. Accordingly, judgments were made by faculty at field research sites as to which competences were relevant to which subareas in each of three selected occupational programs. The work competences were presented in the form of illustrative cases.

Stage 3
(Validity)

Study C3: What is the concurrent validity of the documentation of work competence?

A basic element of the Work Assessment Model (Working Paper No. 8) is the inventory of work competences. If the inventory is valid, there should be a high degree of agreement between the judgments of students and their work supervisors as to which competences were actually learned on a particular job. This study examined that question by locating supervisors and asking them to fill out inventories independently as they applied to individual students; i.e., to provide an indication of the degree of blind verification.

General Procedure and Schedule

The overall plan and procedures for carrying out the validation studies were dictated by the general purposes of the work. It was clear that the studies must be carefully designed and centrally coordinated but must involve the participation of faculty and students at a number of institutions. Accordingly, the work was carried out by ETS staff in cooperation with 24 institutions throughout the country (see list of participants at the front of this report; see Appendix 1c for identification of those areas in which individual institutions participated). Coordinators at these institutions collected the substantial amount of necessary information from students and faculty according to study designs and data-gathering procedures approved by the CAEL Steering Committee. The detailed procedures are described in Chapters IV through VI and the appendices noted there. The following paragraphs outline briefly the overall research plan. It fell naturally into four phases (see Appendix 2c for a more detailed schedule).

July-October 1975: Planning. Whereas this aspect of the CAEL project started formally only in July 1975, initial planning was discussed with the Steering Committee in the previous January and design of the studies was actively underway in the spring. The staff sought advice from committees and consultants as materials were developed during the summer, and the studies were still being redesigned as late as October on the basis of discussions with institutional site coordinators. Another important aspect of planning was the selection of field sites and orientation of institutional participants. The 24 participating institutions were selected by an external committee in late July. To help defray expenses institutions received $2000 to $3500 depending upon the number of areas in which each chose to participate. At the institution's option, some faculty and students were reimbursed for participation. The first orientation meeting was held at the October, 1975 CAEL Assembly meeting in Philadelphia at which time participants reviewed the research plan and research materials. See Appendix 2e for additional general information provided for orientation of field site participants.

November 1975-March 1976: Data Collection. Revised materials for the studies (Appendix 4) were shipped to participants in November and data collection proceeded over the next four months. As is usually the case with complex field studies, it was anticipated that various problems would arise in midstream and that modifications would be necessary. Consequently a meeting of study staff and institutional participants was scheduled to be held in Chicago in mid-January (see Appendix 2f for agenda). Each study was reviewed in some detail; as a result, numerous minor and several major mid-course corrections were made. (See Field Research Progress Report for a summary of the meeting.) In particular, one study was discontinued because the design did not adequately fit the programs of the participating institutions; also, special efforts were undertaken to obtain additional information about the characteristics of individual programs and circumstances that might condition the conclusion or generality of results.

April-June 1976: Data Analyses and Interpretation. Data analyses (described in Appendix 5) was delayed one month in order to include as much late arriving data as possible. Most available data were reported in an initial description of "tentative results" reported in late May. This report of tentative results to all participants and others familiar with the studies was part of a general plan to solicit advice widely regarding any misinterpretations or errors in the data (see Appendix 2h). As expected, careful checking by institutional participants revealed several instances of data incorrectly reported. Also systematic checks of all computer analyses at ETS revealed several errors to be corrected.

July-September 1976: Report Preparation and Review. The first draft of the complete validation report was finished in early August and reviewed by the same fifty-odd people who had earlier received the report of tentative results. The full report was reviewed and approved with revisions by the Executive Committee at its meeting on September 16-17, 1976.

Interpreting the Results

Since the body of this report is intended for a broad audience, technical aspects of the design and data analysis are relegated to the appendix. Nonetheless, there are some characteristics of the data and some basic aspects of the analyses that are important to understand at the outset. Consequently, this section describes very briefly the general approaches to examining the data collected, some factors affecting correlations between judgments, and some qualifications that are important to bear in mind when attempting to draw implications from these findings.

All of the data in this field research represent judgments of one sort or another, usually faculty judgments. There are judgments as to whether learning is creditable, how much credit should be awarded, what level of learning is exhibited, and so on. Different sets of data bear upon the validity, the reliability, or the value of assessment to the student, but questions put to these data always concern the extent of agreement among judgments. There are two types of questions: 1) questions concerning differences among groups (e.g., departments, institutions, disciplines), and 2) questions concerning the consistency among raters within a single group.

These two types of questions have quite different educational implications. Faculty judgment in two institutions may differ on the amount of credit a particular instance of learning deserves and the difference may be justified. Different treatment of the same learning may be dictated by different programs, degrees, or educational philosophy. On the other hand, faculty within a particular unit are presumably working on the basis of the same or at least comparable procedures, policies, and guidelines. In this latter case a reasonable degree of consistency among judgments is expected as evidence of commonly viewed objectives and an equitably administered program. With each set of data it is legitimate to ask, "To what extent should there be faculty agreement within and across institutions?"

The two types of questions also call for different types of analysis. In analyzing differences across groups the differences are easy to describe; an average or percentage for each group usually suffices quite well. The problem is rather to determine whether the differences are real, or perhaps no greater than might be expected merely on the basis of chance fluctuations. In this case the generally applicable statistical procedure is analysis of variance (ANOVA). In analyzing faculty agreement within groups the problem is to describe the degree of agreement in a way that is useful and not misleading. In this case the preferred mode is correlational analysis.

These types of analysis represent broad fields of statistics that are closely related. Some applications can be quite complex, though most of the analyses here are not. The following sections indicate roughly the general character of the two types of analysis. Technical aspects of particular applications are described in Appendix 5.

Analyzing Differences Among Groups

There are various tables throughout this report that show averages or percentages for each of several groups. Sometimes the groups are faculty at different institutions, sets of portfolios from different fields, students who have and have not gone through assessment, and so on. The data presented may be average number of credits awarded, percentage of students rated "high," or whatever, depending upon the nature of the study. From an educational or substantive standpoint, each of those tables must be interpreted individually in light of the context and the particular data involved. From a statistical standpoint, however, there are specific questions one is likely to ask. Table III-1 shows some hypothetical data and the sorts of questions usually asked through analysis of variance.

This figure shows hypothetical data on how many students at each of four institutions achieved the "highest level of competence" with respect to some learning outcomes as it was assessed by two different methods. Any table of this sort contains three findings (or questions), each independent of the other. They are associated respectively with the differences between the rows, the differences between the columns, and the interaction of rows and columns. Thus, one can ask:

-- Are the two methods, A and B, really different? Is .52 actually different from .58, or can that discrepancy be attributed to a chance fluctuation? In this case, the answer would likely be chance.

-- Does the proportion rated "high" differ significantly from institution to institution, or can the differences be readily explained as random variations? With these sample sizes and these observed differences, the differences would likely be judged real.

-- Does the relative proportion judged high by the two methods differ significantly from institution to institution? Here again the answer would likely be yes. Though the proportion rated high is about the same for each method at Colleges 1 and 4, there is a large reversal of pattern in the case of Colleges 2 and 3.

These sorts of questions apply to numerous sets of data throughout the report, though obviously the context varies. The purpose of the analysis of variance is to determine whether the apparent differences are

Table III-1

Hypothetical Illustration of Analysis of Differences Across Groups: Proportion of Students in Four Institutions Judged to Have Attained the Highest Level of a Particular Competence, as Assessed by Two Methods

Institution	N	Type of Assessment Method A	 Method B	Average for Both Methods
1	18	.40	.45	.42
2	22	.40	.80	.60
3	15	.56	.34	.45
4	20	.74	.73	.73
Overall	75	.52	.58	.55

likely to be real differences or merely differences of a magnitude that could easily occur by chance. The question of practical or educational significance is a totally different matter. Thus, whether a "statistically" significant difference among institutions in Table III-1 has major or minor practical significance, is a judgment the reader must make. But if an observed difference between two averages is not statistically significant, the two averages should always be considered not different.

Each analysis of variance carried out for the various studies is described in Appendix 5. Where numerous analyses of a particular type were undertaken, an illustrative outcome is presented so that readers so inclined can ascertain the exact method used. In the text these analyses are typically referred to only in identifying instances where differences were or were not statistically significant.

Analyzing Agreement Within Groups

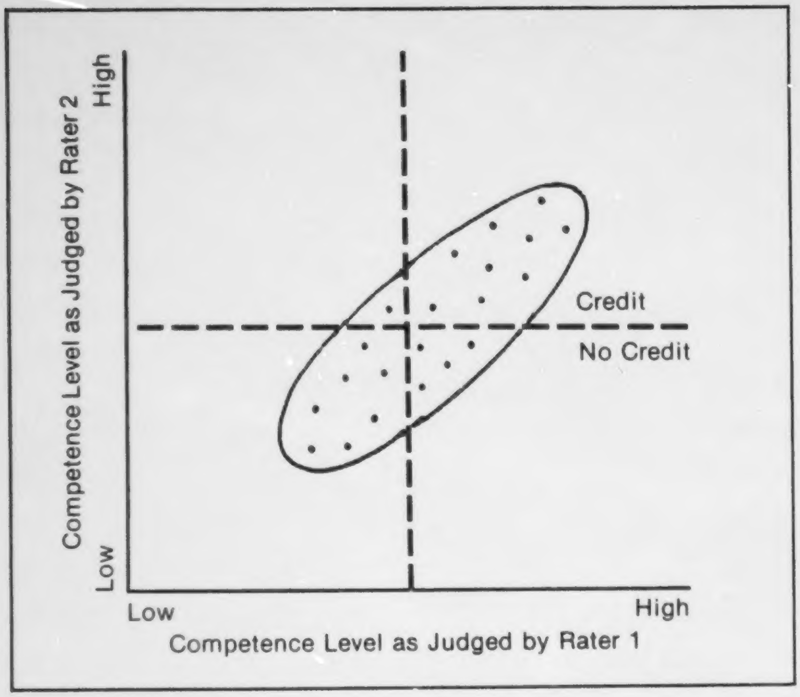
Perhaps the most frequent and important question put to the data reported here is, "To what extent did the faculty raters within the institution or department agree?" That is, how consistent are their judgments? There is a very common tendency to evaluate the degree of consistency between a series of judgments by two people as the percentage of times the two individuals agree. There is a problem with that procedure, as an example can easily illustrate.

If two judges agree 90% of the time as to whether individual students should or should not receive credit, one might conclude superficially that the two judges are in good agreement. But if each judge typically passes 95% of all students, they cannot agree on less than 90% and, in fact, that level of agreement indicates that they disagreed at every possible opportunity concerning who should not pass. Thus, the 90% agreement actually represents a higher level of inconsistency and results merely from the fact that both judges pass a large percentage of students, not from any agreement at all as to which students should receive credit.

Because of this important shortcoming in "percent agreement" as a measure of consistency between raters, the preferred mode of analysis is typically correlational analysis. As an example, Figure III-1 shows schematically the basic idea of describing the relationship between two sets of judgments by means of a correlation coefficient. The top portion of this figure shows the familiar correlation scatterplot of 22 pairs of ratings that might have come from Institution 2 in the previous table. Each dot in the scatterplot represents a pair of ratings for one student. The shape of the swarm indicates that high ratings by one rater are likely to be associated with high ratings by the other rater. There are various ways of computing an estimate of the degree of relationship depicted; i.e., various ways of computing a correlation coefficient that will express the degree of correspondence in the two sets of ratings. When used in this manner, the correlation coefficient is usually considered to vary from .00 (no systematic relationship) to 1.00 (a relationship where all ratings correspond exactly).

The bottom portion of Figure III-1 illustrates one way of summarizing the swarm of dots. The cells of the four-fold table show the count of dots in each quadrant of the swarm. A correlation coefficient is easily computed from these cell figures as follows:

$$R = \frac{(7 \times 10) - (1 \times 4)}{\sqrt{11 \times 11}\ \sqrt{8 \times 14}} = .57 \quad .$$



		Rater 1		
		No Credit	Credit	
Rater 2	Credit	1	10	11
	No Credit	7	4	11
		8	14	22

Figure III-1. Hypothetical correlation scatterplot and corresponding four-fold table showing a correlation of about .60.

It should be noted that there are two ways in which the two raters do not agree. They do not agree on what proportion of students should receive credit (the "rater-error"), and they do not always agree on which students should receive credit (discrepant rankings by two judges). When estimates of interrater reliability take both types of error into account, we refer to the coefficient as "interrater agreement;" when only discrepant rankings of students are considered to be error, the coefficient is here termed "interrater consistency." Both types of error are reflected in the coefficient computed above so it can be considered an estimate of interrater agreement.*

There are many types of correlation coefficients. The particular coefficient calculated above is called a phi coefficient. One of its main advantages is simplicity and adaptability to almost any sort of data that can be sorted into two categories (e.g., high-low) on each of two dimensions (e.g., first and second rating). Phi coefficients are used at several points in this report, but in most instances a more generally useful (intraclass correlation) coefficient is employed. The latter coefficient is more complicated but can be interpreted in a generally similar manner as described above (see Appendix 5 for a description; Tinsley and Weiss, 1975, provide a useful introductory discussion).

For the benefit of readers not familiar with correlational statistics, Figure III-2 shows what correlations of different size look like. When R=.00, those students who fail on a first assessment have a 50-50 chance of passing on a second assessment, and the odds are the

---

*Even though phi is a product-moment coefficient it incorporates the rater-error due to the fact that disproportionate marginal frequencies create an upper bound less than unity.

Figure III-2

Illustration of Correlation Coefficients Ranging From no Relationship (.00) to a Perfect Relationship (1.00)-- Each Based Upon Two Assessments of 200 Students

Second Assessment

		Fail	Pass	
First Assessment	Pass	50	50	100
	Fail	50	50	100
		100	100	200

R = .00

Second Assessment

		Fail	Pass	
First Assessment	Pass	30	70	100
	Fail	70	30	100
		100	100	200

R = .40

Second Assessment

		Fail	Pass	
First Assessment	Pass	15	85	100
	Fail	85	15	100
		100	100	200

R = .70

Second Assessment

		Fail	Pass	
First Assessment	Pass	0	100	100
	Fail	100	0	100
		100	100	200

R = 1.00

same for those who pass the first assessment. Thus, there is no agreement between the two assessments as to who passes and who fails. When R=1.00, every person passed on the first assessment also passes on the second assessment; similarly, every person failed on one assessment fails on the second--perfect agreement. Correlations of .40 and .70 show intermediate tendencies for two assessments to agree on who fails and who passes. There are a number of factors that affect the size of the coefficient and how one interprets it as a measure of reliability of judgments. These factors bear some discussion.

Factors Affecting Reliability of Judgments

In real educational situations perfect or near perfect agreement (a reliability coefficient approaching 1.00) never happens except by chance in a small sample of students or when the judgments are not truly independent; e.g., when a second rater has somehow become aware of and influenced by a prior rating. Carefully constructed standardized examinations typically have reliabilities in the low .90's; i.e., would yield pass/fail decisions somewhere between the R=.70 and R=1.00 depicted in Figure III-2. Consistency among independent judgments is another matter. With the type of judgments represented in this report it is more reasonable to consider coefficients in the range of .70 to be fairly good and a coefficient of .40 to represent modest agreement.

In truth, however, there can be no easy rule of thumb. While subjective judgment of classroom teachers, professionals, and other experts are notoriously unreliable (see Warren, 1971; Coffman, 1972; and Wiggins, 1973, pp. 128-140 for representative data and useful discussion), actual data varies widely depending upon the specifics of particular types of judgments and conditions. Thus, the data presented in this report cannot be cast upon some "good-poor" framework for easy interpretation. What seems most useful is some comment on the most important factors which tend to determine whether expert judgments are consistent (reliable) or not. The following paragraphs identify five considerations worth special attention:

1) Number of raters. It is said that two heads are better than one, and that aphorism has special meaning with respect to the quality control of assessment. As compared to a double rating, single raters are more likely to make mistakes, to interpret policy incorrectly, and to let their judgment be unconsciously colored by a positive or negative "halo" about the student that may be due to some extraneous fact or earlier encounter. It is a firmly established fact in psychometric literature that a composite judgment based upon multiple raters is more reliable than one based upon a single rater. In fact, the relationship can be expressed mathematically as

$$R_n = \frac{nR_1}{1 + (n-1)\,R_1} \quad ,$$

where R_1 is the reliability of one rater, R_n is the reliability for n raters, and n is the number of raters. Table III-2 illustrates this relationship. Knowing the number of ratings on which a reliability coefficient is based, one can see roughly from this table how much the reliability would likely go up or down if the composite rating were based upon more or fewer individual ratings. For example, if the reliability of one judge's rating is .50, the reliability of the combined judgment of two judges is .67. The obvious benefit of the extra judgment must be pitted against the added cost.

Table III-2

Relationship Between the Reliability of a Composite Rating and the Number of Ratings on Which the Composite is Based

Number of Ratings	Size of Reliability Coefficient					
1	.20	.30	.40	.50	.60	.70
2	.33	.46	.57	.67	.75	.82
3	.43	.56	.67	.75	.82	.88
4	.50	.63	.73	.80	.86	.90
5	.56	.68	.77	.83	.88	.92

Table III-3

Improvement in the Reliability of Essay Grading Through Use of Multiple Readers and Multiple Samples of the Student's Writing (from Godshalk, et al., 1966)

	Number of 20-Minute Essays	
	1 Topic	5 Topics
5 Readers	R = .76	R = .92
1 Reader	R = .38	R = .76

Throughout this report studies of the consistency of various judgments are based upon two to five judges. In order to avoid variability in estimates of consistency due merely to variation in the number of judges, all estimates are "corrected" so as to represent the combined judgment of two judges. Such estimates seem more defensible than estimates for one judge or five judges because it seems likely that most decisions regarding credit for experiential learning are based upon a small team judgment or an individual backed up by a reviewer.

2) Sampling of Performance. Just as it is important to get more than one opinion of a student's learning, it is also important to get more than one limited sample of learning; both represent multiple views of performance and give a more dependable estimate of the student's competence. One brief example of a student's learning may well give an atypical impression. Godshalk, et al., (1966) provide extensive evidence on both points. They reported what is probably the most detailed study ever undertaken of essay grading--an especially important type of assessment in crediting experiential learning. Table III-3 shows some concrete data on the importance of obtaining multiple judgments and multiple evidence in order to insure dependable decisions; both substantially improve reliability of the overall judgment. Other evidence presented by the same authors indicates that length of the essay is not in itself so important. Forty-minute essays were hardly more reliable than 20-minute essays, but two 20-minute essays on different topics did provide a significantly more reliable estimate of writing ability than did one such essay.

3) Clarity of criterion standards. It is self-evident that a mathematics test with one right answer for each problem will likely be graded more consistently by two graders than will a political science essay examination. The same principle applies to expert judgment of any learning or performance. The more delineated the criteria and the more specific the standard for successful performance, the more likely judges can agree on whether a student passes or not. Data from the Advanced Placement Program of the College Entrance Examination Board provide an example. Through many years of experience, the AP Program has achieved high reliability in grading various types of essay examinations and products of student work. For example, during the past three years graders have achieved the following reliabilities in judging essays in art history and actual examples of student art (Chen, 1975):

	Essays in Art History		Examples of Student Art	
Year	Number of Students	Reliability	Number of Students	Reliability
1973	270	.78	252	.73
1974	384	.79	349	.78
1975	436	.85	380	.80

These reliabilities have been attained in large measure because of close attention to criterion standards and training of faculty "readers" to interpret those standards appropriately. Standards for each examination are defined on a point scale (e.g., 1-9) with respect to the character and quality of the performance expected. Also each point on the scale is illustrated with an actual example that shows in concrete terms what constitutes good, poor, and indifferent performance. Finally, readers are trained through practice sessions to recognize consistently student work that does and does not meet the stated standards. Smith (1975) has described this process in an informative account. It dramatizes the importance of judges thoroughly understanding the criteria of judgment if fair assessments are to result.

4) Range of competence. Reliability or agreement among judges is a matter of consistent discrimination among students, portfolios, or whatever is being judged. Consequently, if there is very little difference among the items or people being evaluated, there is less likelihood of consistent discrimination among them. This can happen when the sample of objects to be evaluated is homogeneous because of pre-selection or restriction to a special group of high level candidates. For example, Scott (1974) and McReynolds (1971) report somewhat low single rater reliabilities of .26 and .45 (.41 and .62 for two raters) in the judging of professional research papers by highly qualified experts.

5) Common training and point of view. Perhaps it is too obvious to require mention, but judges are not likely to agree if they have made little effort to ascertain on what common bases they are to make judgments. Earlier two general types of rating errors were distinguished: a) rater-error or the tendency for one judge to persistently rate much higher or lower than another, and b) inconsistency, the tendency for two raters to give different rankings to the individuals or objects being judged. There is every reason to believe that the former is much easier to correct than the latter through development of guidelines and training of judges.

A liberal or conservative rater-error shows up dramatically in differences among faculty in different institutions; in which case it is perhaps more aptly called a different "point of view." The second type of error--inability of two judges to agree in discriminating which learner has demonstrated acceptable performance--is probably less subject to easy correction and requires careful specification of standards and effective training of judges as to what learning is being assessed and how to recognize good and poor instances of it. In estimating reliability of judgment we have referred here to "interrater agreement" when the coefficient takes both types of error into account and "interrater consistency" when only differences in rankings are included as error (p. 19). Since the latter coefficient takes account of only one type of error, it is almost always higher. The former coefficient represents the best estimate of actual reliability but the latter coefficient represents the level of rater reliability that might be attained if discrepancies in the average rating assigned by different judges can be minimized through training of judges.

Some Qualifications

It is useful to bear in mind some important considerations that may condition or qualify interpretation of specific findings reported in subsequent chapters. There are several predictable characteristics of most field research that might be viewed as shortcomings or problems. For example: 1) the participating institutions are typically very different; 2) the research materials and procedures often do not match exactly the educational programs at those institutions; 3) the research results will likely vary from institution to institution; and 4) results do not necessarily apply to other institutions. It is useful to consider these generalizations briefly. They are associated with certain strengths and weaknesses of field research and each has implications with respect to how the studies are carried out and how the results are interpreted.

1) Participating institutions are very different. For the overall purpose of the validation studies, variation among the participating institutions is a strength because such diversity makes the results more generalizable to a diverse population of higher institutions. On the other hand, diversity may be a weakness of research design in the case of particular institutions where for one reason or another the research design or materials seems inapplicable. Because of institutional variations that cannot always be anticipated, field research usually requires adjustment in research procedures. Such was the case here, but adjustments within an overall procedure do not necessarily make the results fully applicable to particular institutions. Feedback was solicited from participating institutions concerning any discrepancies that would lead one to question data from a particular institution or to hedge any conclusions reached. Several sets of data were discarded or qualified on that basis.

2) Materials and procedures do not exactly match program practices in individual institutions. In some types of studies this condition is inevitable due to the fact that research must employ reasonably standard (illustrative) materials and procedures across institutions so that results can be interpreted across institutions in relation to some known quantity. That comparability of research design is, of course, a strength; but a weakness may lie in the fact that the standard research procedure may depart too much from the reality of operating programs or that individual institutions cannot compare results with their own specific practices as directly as they might wish. It is important to recognize, however, that this type of study is concerned with the evaluation of principles and

assumptions, not with any implied judgment concerning the adequacy of specific materials or procedures. Therefore, it is not critical that illustrative portfolios in Study B1, for example, be complete or ideal in all respects or that they represent exactly the assessment process at a given institution. That same study includes a second phase in which portfolios are judged using materials from actual local students and the particular assessment procedures prescribed by each institution. In this second phase the study procedures match program practices, but comparability across institutions is lost and interpretation of results is consequently restricted in a different way. Thus, the two phases using illustrative and actual materials and procedures offer complementary advantages and information.

3) Results will vary from institution to institution. Variation in results from institution to institution is to be expected and can even be useful in a sense. A pluralistic system of higher education will show systematic differences among institutions in matters concerning what is creditable, how much credit is awarded under particular circumstances, etc. Field studies that document such differences will help individual institutions get a sense of that variation so that they can better evaluate their own practices. In this connection, one primary reason for using illustrative material in some studies is to provide results that are directly comparable from one institution to another. Thus, in some analyses of faculty judgment one would expect to find some differences among colleges concerning what they are willing to credit but hope to find reasonable consistency within colleges.

There is also a weakness in the fact that results vary from institution to institution. That is the danger of unwarranted conclusions regarding results for single institutions. Data from one institution may be misleading either because the institution is different, the study was carried out differently at that institution, or simply because of random variation which is to be expected. To guard against such misinterpretations, results for individual institutions are listed with no identification. Furthermore, readers should be very cautious in interpreting data based upon small samples. Such results are quite unstable.

4) Results do not necessarily apply to other faculty or institutions. In interpreting research results that may appear quite clear, it is always important to consider how generalizable the findings actually are. In one sense it is obvious that one cannot be sure that conclusions apply to other situations, but if there is variety in the sample, it is more likely that the findings are general. Clearly the institutions involved are not, in any sense, a representative sample of higher education nor are the students and faculty who participated. The reader must judge in what respects the samples here can be assumed generally similar to large populations of students, faculty, and institutions. While the institutions and individuals are clearly a select group regarding interests and experience, they are probably not greatly different in such basic qualities as ability to judge competence when they see it.

Finally, there is an inevitable lack of control in any field research. Those designing the research and reporting the results cannot be thoroughly familiar with local conditions that affect results; site supervisors cannot be thoroughly familiar with all the ways in which seemingly unimportant details or unwitting deviations from procedure can distort data. Studies of judgment are especially troublesome. It is easy for one judge to bias the judgment of another without knowing it; it is also easy for procedural variations or inappropriate materials to create error in judgments. The former normally increases observed agreement; the latter is more likely to decrease observed agreement. It is always uncertain how these balance out, and the reader must make his or her own judgment from the available evidence.

Chapter IV

STUDIES OF ASSESSMENT OF INTERPERSONAL SKILL

Thomas F. Donlon and Richard R. Reilly

The Priority Areas established by the CAEL Steering Committee in 1974 purposefully reflected a balanced diversity of characteristics. The designation of the assessment of interpersonal skill as one of four priority areas recognized the need to demonstrate the assessment of a complex skill. Further, it responded to the many experiential programs which emphasize growth in interpersonal competence. Finally, it involved CAEL in activity which would draw upon and apply assessment methodology from a number of applied areas.

The stimulus of Working Papers Nos. 4 and 5 established the framework within which the validation studies were undertaken. As stated in Chapter III, it is clear "that CAEL is not producing any...standardized tests." Rather CAEL is "fostering certain assessment procedures." Working Papers Nos. 4 and 5 in responding to the priority designation, described a variety of methods for assessment, and reviewed them in the context of institutional considerations and the need for an underlying theory or rationale. As the authors remark:

> "The area of interpersonal skills is both challenging and complex. The work we have attempted is like that of the other CAEL Priority Areas, but it is also quite different. With them, we have focused on problems of assessment, broadly considered...But we differ from them in the substantive content or focus of our area. Whereas the other areas are considering the assessment process as it might relate generally to any content area, ours has been centered on a specific set of competences."

The Working Papers Nos. 4 and 5 focused on 11 assessment methods, recommending their organization into a core program of five basic methods: 1) checklists, 2) experience summaries, 3) interviews, 4) validation documents, and 5) ratings. These were defined on both logical and practical grounds. Attention was called to six other useful approaches: 1) topical essays, 2) paper-and-pencil tests, 3) leaderless group discussions, 4) simulations, 5) work samples, and 6) content analysis.

The analysis of methods in this area is made difficult by their complexity. The distinctions among methods which are suggested by the recommendations are recognized as imperfect. Nonetheless, there is a sizable number of approaches, and only a smaller subgroup of them could be studied in the validation phase. The three studies in this chapter, in terms of the content of their methods may be described as follows:

A1 - uses ratings and a portfolio structure that documented accomplishment

A2 - studies the interview and the role-playing simulation

A3 - uses ratings and, to a limited extent, a content analysis of self-reported judgments of interpersonal competence.

The field research derived from the Working Papers in a number of ways. Thus, the model for the interview in Study A2 was the sample structured interview in CAEL Working Paper No. 4. The student-centered review of assessment, Working Paper No. 5, was an integral part of the assessment process studied in A3. The impact of the Working Papers was also reflected in the selection of methods for study; in general, it is higher-priority methods that are being used.

The studies are not narrowly concerned with validating explicit material drawn from the Working Papers. There was a need to permit a number of adaptations of the Working Papers to the specific contexts of individual institutions. A number of other considerations necessarily governed the design of the studies, including the need to cover both reliability and validity. The Working Papers, then, are best conceived as background and stimuli for the validation projects, rather than as the focus.

It is appropriate to establish a perspective for the CAEL Area A validation within the context of the present measurement technology in this area. Increasingly in recent years education and psychology have recognized the development of interpersonal competence as a desirable goal for the fully mature individual. But is is difficult to generate a fully satisfactory definition of interpersonal behavior. We do know that it is extraordinarily sensitive to environmental influences, particularly the social environment. Yet most measurement strategies must seek to manipulate the social environment. In so doing, they may distort the very behavior they seek to observe. These problems of definition and of distortion pose significant difficulty for successful measurement. The perspective for evaluating what is done must not proceed from ideals but from a view of what is practical and possible. The long history of the assessment of interpersonal competence in occupational settings, as presented in the literature of military and industrial psychology, reveals that, while much can be done, the best efforts fall far short of the ideal.

This perspective should, of course, not be applied narrowly to CAEL. The need for opportunity for growth in interpersonal competence will be an increasing component of education in the years ahead. Faculty, students, and administrators will confront precisely these problems. The work is important, and it would be unfortunate if it were abandoned. The perspective presented here is directed as much to the area of education for interpersonal competence as it is to the narrower scope of the CAEL project.

The data reported in these studies were secured with the cooperation of the following institutions:

Alverno College
Hartwick College
Macalester College
Metropolitan State University
San Francisco State University
Staten Island Community College
Stephens College
University of Akron
University of California at Los Angeles
University of Oregon

The efforts of a sizable number of persons are reflected in the results.

Study A1 - TO WHAT EXTENT CAN FACULTY DIFFERENTIATE VARIOUS INTERPERSONAL COMPETENCES IN TERMS OF THEIR RELEVANCE TO DEGREE PROGRAMS IN DIFFERENT SUBJECT MATTER AREAS?

The CAEL Priority designations reflect strong interests on the part of major groups in the field of experiential education. The potential value, then, of assessing interpersonal skills is established. Institutions and individual learners alike will be seeking to establish creditable competence in this area.

The establishing of credit will vary, in process, from situation to situation. A common theme, however, will be the role of faculty in integrating the complex body of information that is the record of student performance. The present study is an examination of a number of characteristics of faculty performance as an initial step toward understanding the potential for useful assessment processes in the interpersonal field.

Description of Study

One indication that there exists a clear and useful rationale for crediting interpersonal skills is the demonstration of agreement among faculty with respect to which interpersonal competences are relevant for various degree programs. The basic purpose of this study is to examine the extent to which such faculty agreements are demonstrated in the context of three areas of study: management, education, and psychology.

At each cooperating institution a number of judges were selected (optimally three but sometimes fewer) for each of the three fields, each judge was asked to independently evaluate five illustrative cases. Each illustrative case consisted of a set of documents relating to a given student's request for credit based upon attainment in interpersonal competency. The specific contents of the documents varied from one illustrative case to another, but typically the set contained 15 to 20 major pieces of information. These included institutional degree pacts in experimental programs, statements of specific areas in which competence is claimed (e.g., public speaking) and appraisals, letters, institutional ratings developed in the course of presenting a claim for competence. These documents were adapted from actual student cases encountered at cooperating institutions.

The illustrative cases selected were appropriate to a specific field, so that there were 15 cases in all: 5 for psychology, 5 for education, and 5 for management. The evaluations essentially consisted of appraisals of the amount of credit which might be awarded to the illustrative case in question. To insure a balanced appraisal, judges were asked to consider each of three levels: lower division, upper division, and graduate. For the general analysis of rater consistency, however, distinctions among levels were disregarded, and the basic measure used is the total amount of credit, combining all three levels, that a particular judge awarded to a particular illustrative case.

Institutional Context

There were a number of minor differences in the organization of the study from one campus to another. In general, the rating form suggested by ETS for use in conveying faculty estimates of credit was used without change. The faculty who participated tended to have some connection with an experiential program on the campus. At San Francisco State University, for example, six of the ten faculty judges were members of the "CEEL" (Credit by Evaluation for Experiential Learning) project at that institution. At Alverno, nine faculty members had an average of 2.7 years experience in "nontraditional higher education." At Stephens College, the faculty description observed that "none have extensive experience in nontraditional education." The range of faculty experience was quite large. The faculty raters tended to report some discomfort in rating interpersonal competence without an opportunity to see the student. As described in an institutional report, "The greatest difficulty lay in trying to set aside a fundamental attitude that looks for demonstration rather than documentation of prior learning." Said another, "We found ourselves evaluating experiences, and not the interpersonal skills which were the outcomes of these experiences."

These reports bear heavily on the interpretation of Study A1. It is clear that the structure of the information which is presented to faculty is an important determiner of their performance. The present effort could not consider the many dimensions of this structure, could not vary it systematically. The field reports indicate that functional programs at institutions would not make an appraisal of interpersonal skills which was totally portfolio-based.

Institutions differed in the extent to which there were current procedures for securing credit for interpersonal skills based on prior learning. At one institution,"there is no allowance for awarding credit for previous experience." At another, which credited a variety of prior and sponsored learning experiences, "interpersonal skills are included in degree programs." At a third, "Interpersonal skills constitute one of the eight competences required..." At this institution the task of Al was counter to the procedures which were ordinarily followed.

Such variation was expected and is a necessary aspect of field research. While it makes it difficult to conduct systematic experiments, it increases the successful generalization of any findings which do emerge, by removing from them the fragile quality associated with laboratory findings gathered under highly specialized conditions; but we do not assume that any experience is equivalent to learning--even academic experience.

Results

It may be helpful to begin by reviewing the ways in which faculty within a department agree and disagree in awarding credit for illustrative cases of interpersonal competence acquired through prior learning. Table Al-1 illustrates the basic data of this study for one department at one institution. It is apparent in this table that the five illustrative cases are awarded different amounts of credit: Case 3 receives the highest awarded amounts for each judge and overall, while Case 5 is consistently given a smaller amount of credit.

But there is evidence of a number of sources of error or inconsistency. For example, faculty member Z tends to award somewhat more credit, on the average, than does X or Y. Accordingly, how much credit a student receives for interpersonal competence may depend upon the general orientation of the faculty member. Secondly, it is evident that faculty members do not agree exactly on individual cases. The

Table A1-1

Credit Awarded for Experientially Acquired Interpersonal Competence Relevant to Psychology Requirements in One Institution -- Five Illustrative Cases Rated by Three Faculty Members*

Illustrative Case	Credit Awarded by Psychology Faculty Member			Average for Three Ratings
	x	y	z	
1	1	10	10	7.0
2	6	6	16	9.3
3	18	14	18	16.7
4	8	3	7	6.0
5	5	2	4	3.7
Average for 5 Cases	7.6	7.0	11.0	

---

*Reliability (agreement among faculty) for these ratings was .75.

most obvious discrepancies are the amount of credit estimated for Case 1 by Judge X and the amount estimated for Case 2 by Judge Z. It is possible to summarize the degree of consistency in such a table and calculate a reliability coefficient for the average credit that a case might receive from three raters. For Table A1-1, this coefficient would be .75, which is quite good for judgments of this type.

Table A1-1 indicates that the average amount of credit awarded at the institution across all five cases and all three judges was 8.5. This number is a useful summary index of the tendency to win credit in psychology at this institution. It combines, in a sense, the 15 individual experiences of the table into a single value. The following questions focus on the principal results of analyses of this type.

Question: 1 Are there differences among institutions in the extent to which psychology faculty award credit for the same illustrative cases of interpersonal competence? Management faculty? Education faculty?

Summary values such as the means in Table A1-1 are helpful in considering the three related components of Question 1, which focus on comparisons of the different institutions as they award credit in the three content areas. Table A1-2 presents the summary values in each of the content areas for each of nine institutions together with the overall means.

In psychology there is evidence of a considerable variation in tendency to award credit, ranging from 3.2 for Institution G to 30.8 for Institution E. These two institutions are the most extreme, with the remaining schools ranging from approximately 5 to 15. In Management there is a similar pattern, ranging from 21.3 for Institution F to 2.7 for G. Again, the range is sharply reduced for the remaining institutions. Finally, in Education the span is from Institution E, with 7.0, to Institution G, with 49.2. For each of these three content areas: Psychology, Management, and Education, the differences among the means for the institutions are statistically significant, indicating that in each field the views of the faculty teams differ from one institution to another.

An analysis of variance table, related to the content of Table A1-2, is presented in Appendix 5a, Analysis 1.

Question: 2 Does the extent to which different departments award credit for such cases vary from one institution to another? (e.g., in one college the psychology faculty awards credit generously, while in another college it may be the education faculty.)

Question 1 is essentially answered by considering the entries to each subject content column in Table A1-2. The present question is best answered by considering the "All Fields" column and the patterns within individual institutions (rows of the table). Entries in the "All Fields" column are the sum of judgments across the content fields. As the "All Fields" column indicates, there is general variation in the level of credit awarded, irrespective of field. Further, this general level, being summative, provides slightly different information from the separate subject content columns. For example, Institution D in general evaluated the portfolios most severely, awarding the least credit overall although D is not the most severe institution in all of the areas.

There is marked variation in the level of credit awarded within field in these institutions: Education faculty members ascribed more credit than Psychology faculty, with Management faculty third. These

Table A1-2

Average Amount of Credit (Semester Hours) Awarded Illustrative Cases for Interpersonal Competency Gained Through Prior Experiential Learning

Institution	Psychology*	Management*	Education*	All Fields
A	7.9	7.7	9.3	8.3
B	21.9	8.2	16.5	15.5
C	15.3	5.8	14.7	11.9
D	5.2	5.2	7.2	5.9
E	7.7	5.7	14.3	9.2
G	34.6	5.6	7.3	15.8
H	14.7	21.3	41.8	25.9
I	3.4	2.7	49.2	18.4
J	11.9	4.9	19.9	12.2
All Institutions	15.3	8.4	22.5	15.4

*Each entry in these columns is based upon two or three ratings for each of five cases. The five cases were different for the three fields.

differences cannot be interpreted as clear differences among fields, however, for the samples of cases were rather small and the nature of the illustrative cases varied from field to field.

Though it is difficult to draw a clear conclusion regarding differences among the three fields across institutions, it is possible to examine whether differences between fields are greater in some institutions than in others. To what extent are the differences among institutions related to subject matter? In general, the differences among institutions seen in the "All Fields" column is repeated to some extent within subject fields, but there are obvious variations from institution to institution.

When these variations are summarized, there is statistically meaningful evidence that some institutions are unusually stringent (or flexible) in awarding credit in Psychology, while other institutions are unusual in their pattern of credit in Management or in Education. This is further evidence of the variations among decision standards which are found within institutions as well as among them.

An analysis of variance table relating to Question 2 is presented in Appendix 5a, Analysis 2.

Question: 3 On an overall basis, to what extent do faculty tend to agree when awarding credit for these illustrative cases of interpersonal competence? How consistent (reliable) are credit awards within individual psychology departments? In management? In education?

The appraisal of rater consistence as reliability may be approached in a number of ways. One approach is to look at the actual discrepancies in pairs of judgments. Each time that two raters judged a given illustrative case, a potential for difference was created. For two raters there was only one such possible difference, while for three raters there were three possible differences (Rater 1 from Rater 2, Rater 1 from Rater 3, and Rater 2 from Rater 3). With five cases at nine institutions, there would maximally be 5 x 9 x 3 = 135 differences in each subject matter column. Table A1-3 presents the distribution of the absolute difference of credit across raters within the schools. Due to the loss of third raters, there were about 100 differences in each distribution.

In general, the preponderance of faculty disagreements were in the ranges less than 9. Thus, in the overall column, disagreements less than 9 account for 68.5% of the observations. Separately, by field, the comparable percentage is 73.9 for Psychology, 78.2 for Management, and 54.6 for Education. About one comparison in eight differs by more than 24 points in credit, on the average. This varies from about one in nine for Psychology to one in twenty and one in five for Management and Education, respectively. Thus, by this index of rater consistence Education raters approached the cases most variably, Management raters least variably.

Table A1-4 presents the distributions of reliability coefficients within the fields. Essentially, those coefficients reflect variation in ratings which seem to be random or unexplainable. The greater the unexplainable variation, the less reliable the ratings.

By this index there is evidence of a number of different levels of reliability in the various institutions and fields. While in some instances the values of the coefficients were sufficiently large to constitute statistically significant results, in others the coefficients were numerically close to zero. Such variation indicates that combining results may obscure the outcomes. Accordingly, no overall summary

Table A1-3

Percentage of Paired Faculty Judgments Differing by Specific Numbers of Credits for Each Field and Overall

Difference in Credits Awarded	Psychology N=109*	Management N=94*	Education N=108*	Overall N=311*
0 - 1	23.9	16.0	13.0	17.7
1+ - 3	20.1	30.8	13.9	21.2
3+ - 6	20.7	22.3	13.4	18.7
6+ - 9	9.2	9.1	14.3	10.9
9+ - 12	9.1	7.4	12.5	9.8
12+ - 15	2.3	3.8	5.6	3.9
15+ - 18	1.9	2.1	5.1	2.7
18+ - 21	1.9	1.6	1.8	1.8
21+ - 24	.9	2.1	1.0	1.2
24+	11.0	4.8	19.4	12.1

*N = number of pairs of faculty judgments reflected in each column.

reliability is presented. However, the general level of reliability within a field may be estimated by considering the distribution of the coefficients. Those for psychology tend to show somewhat greater values than those for management or education. A general picture of limited reliability is presented.

The overall data for Study A1 suggests the following observations:

1) The level of variability among faculty within an institution and a department in awarding credit is quite large.
2) The differences among institutions in awarding credit are statistically significant.
3) The differences among institutions in awarding credit vary from one field to another. Some institutions seem more likely to grant credit for psychology, some for education, and some for management.

It should be pointed out that the sizes of the various samples upon which the values in the various tables in Study A1 are based are quite small. This factor is of considerable importance in discussing particular entries in Table A1-4, for the reliability estimates reported there are reflections of very limited data; it was the rationale for characterizing the distributional properties of these intra-institutional coefficients without associating them with specific institutions. The central tendency of the values, their average, is not affected by the small size, and the conclusion that, overall there is little reliability, is established. There are, accordingly, a number of constraints on the study. The variety of illustrative cases and the number of raters within institutions are examples of factors which reflect the practical limits on the scope of this research.

Table A1-4

Distributions of Reliability Coefficients by Field for the Average Faculty Rating in Awarding Credit for the Illustrative Cases

Reliability More Than	 Less Than	Psychology	Management	Education	Overall
.90	1.00				
.80	.90	2			2
.70	.80	1			1
.60	.70	1		1	2
.50	.60	1		1	2
.40	.50	1	1		2
.30	.40		1	1	2
.20	.30				
.10	.20	1	1		2
.00	.10	2	6	6	14

Study A2 - TO WHAT EXTENT DO DIFFERENT ASSESSORS AGREE ON THE LEVEL OF INTERPERSONAL COMPETENCE STUDENTS HAVE ACQUIRED AND TO WHAT EXTENT IS THERE AGREEMENT BETWEEN DIFFERENT METHODS OF ASSESSMENT?

Study A2 differs from A1 and A3 in its behavioral focus. This effort to appraise actual interpersonal behavior is, of course, consistent with the philosophy of CAEL institutions themselves. There are frequent expressions of the need to anchor the assessment of interpersonal skills in student performance. Further, A2 explores a fundamental assumption of Working Papers Nos. 4 and 5, that assessment will best proceed from multiple strategies or techniques. Behind this assumption lies a broader concept of the need for balance in the relatedness and independence of the various methods which are used.

Description of Study

A number of techniques are available for appraising interpersonal skill through behavioral samples. Two of these methods were selected for examination in this study: the interview and a role-playing simulation. Several important questions about the outcomes of such method comparisons may be studied. Clearly, if assessment is to be practical, it must offer a variety of methods and these must provide a reasonable consensus. Interpersonal skill is a complex and somewhat abstract concept. No single appraisal technique will be useful in all contexts, and in general the appraisal of the assessment processes themselves will be benefitted by a study of the relationships among them.

Each institution was asked to secure 15 students to participate in each of the two appraisal situations: interview and role-playing simulation. The content for the interpersonal situation was that of a counseling relationship, which was broadly defined to include any basically helping relationship in a variety of more specific settings. Thus, there could be reasonable variation in the problem focus of the counseling, from drug addiction to vocational or marital adjustment.

Institutions were invited to base their structured interview on the sample provided in CAEL Working Paper No. 4, adapted as appropriate to the individual case. The role-playing situation was a counseling simulation, also adapted in ways which were meaningful and appropriate to each participating student. The guideline material for the structured interview is presented as Appendix 4e, while the suggestions for the role-playing exercise are given as Appendix 4f.

Both interview and role-playing were appraised by two observers so that within an institution there were as many as 60 observations. The ratings were on a five-point scale from 1 to 5. In all, seven aspects or facets of counseling competence were delineated, and each student was rated on these seven traits and on a summary, overall basis. The rating scale which was used was related to levels of student progress, from (5), skills equivalent to those of a practicing professional, to (1), little or no evidence of interpersonal competence. The faculty rating form is presented as Appendix 4g.

Institutional Context

Data were received from ten institutions. Some of the data were received too late for inclusion in certain of the analyses. In a few cases, not all of the studies could be completed at each institution.

Adaptation to the institutional setting was a primary requirement for Study A2. The student participants varied from one school to another. While counseling interests were predominant, both in students and in faculty, nursing students were the participants at one institution, a mixture of social science majors at another. In one case, the students studied in A2 were drawn from a University Without Walls program at a state correctional center for women.

These institutional specifics were reflected in the content of both the interview and the role-playing exercise. The structured interview from CAEL Working Paper No. 4 was accepted at most participating institutions as the base for the interview assessment, but it was freely modified. Many institutions reported that individual questions were not appropriate in their context and that they had been eliminated.

The study of contrasting methods for assessing interpersonal competence seemed to be well received on these campuses. The effort to evaluate behavioral samples generated considerable interest. There were several expressions of interest in continuing the momentum generated by the study, and making both interview and role-playing regular components of the assessment process. Certain institutions, of course, already had assessment processes which used these techniques. Students were volunteers, with some institutions compensating students for participation, some not. The studies necessarily generated rather complex scheduling problems, requiring as they did the coordination of larger numbers of people than Studies A1 or A3.

A number of institutions described problems with the rating scale for Study A2, which was anchored in levels of education rather than in critical levels of behavior. For some institutions, where graduate programs were too remote to be applicable, there were redefinitions of the five rating categories into general levels of excellence. While the utility of the suggested scale (presented as Appendix 4g) was recognized, in that it offers a kind of common denominator across schools, the correlation between interpersonal competence and level of education was not sufficient for many users to be fully comfortable with it.

Data Analysis

The data analysis for Study A2 focuses on three main properties of the ratings. First, there is the basic property of reliability. Considered separately, do the ratings given by faculty show consistency? If a student is appraised by one rater by the interview technique, is there evidence that this appraisal would be close to that given by another rater? The methods for determining these questions use correlational techniques, studying the way ratings vary together.

Second, there is the basic property of intermethod consistency. Do the assignments by interview and by role-playing agree with each other? Again, the basic analytic technique for answering this question is a correlation between methods as observed for a sample of students.

A third principal property is really a set of properties. It concerns the factors which lead to variation in ratings. To what factors and combinations of factors are they sensitive; which ones seem to influence rating? The formal data analysis considers a number of influences on the ratings, testing the variation in level of rating at the several institutions, for the methods, and for combinations of the two. Much of the description of this formal analysis is reserved for

Appendix 5. A very useful technique for this kind of question is the analysis of variance, and Appendix 5 presents a number of analysis of variance tables relating to Study A2. Exhibits of mean values for various factors and combinations of factors are presented in Tables A2-2 and A2-4 of this chapter.

Question: 4 Do students consistently receive higher ratings for interpersonal competence in a counseling situation when they participate in interviews, as opposed to role-playing? Or are the ratings in role-playing higher? How do these comparisons vary from institution to institution?

Table A2-1 provides an example of the basic 60 elements of data which were available from each institution. It shows the range of variation which is observed across the 15 students. In the example in Table A2-1, there were no awards of "5," the highest level of competence, to any of the 15 students. There were two awards of "1," low levels of competence. Rating systems need to encourage raters to use the extreme values of scales if they apply, for there is a natural tendency to "reserve" them for some anticipated but hypothetical cases. Good competency scales should be well anchored to levels of development within training programs, but it is possible that an extreme value may seldom be relevant (e.g., "published a book" as an instance of communication skill).

A number of analyses are possible in such data. The average rating awarded for interview can be contrasted with that awarded for role-playing. At this institution, the interview average was slightly higher than the role-playing average. It is also possible to correlate the ratings for the two appraisals by a similar method and announce reliabilities or agreement between judges. For the interview approach

Table A2-1

Overall Ratings of Counseling Competence for 15 Students in a Single Institution Assessed by Judges Using Two Different Modes of Assessment

	Ratings of Overall Effectiveness			
	Interview		Role Playing	
Student	Judge 1	Judge 2	Judge 3	Judge 4
1	2	2	2	1
2	2	2	2	2
3	4	2	2	2
4	4	3	4	3
5	2	3	2	2
6	2	2	2	3
7	2	1	2	3
8	2	4	2	3
9	3	2	2	2
10	4	3	4	4
11	3	4	2	2
12	4	4	3	3
13	3	3	2	2
14	3	3	3	3
15	3	2	2	2
Average Overall Rating for 15 Students	2.8		2.4	

Reliability for interview ratings was .46, reliability for role playing was .81. The correlation between interview and role-playing ratings was .53.

reliability was .46, for the role-playing method it was .81. The correlation between the two methods was .53. This correlation should probably be interpreted in the light of the relatively low reliability for the interview. If both procedures could be made more reliable, they would correlate more highly.

A table such as A2-1 could be presented for each institution. Table A2-2 presents nine sets of summary means which are derived from the bottom line of tables such as A2-1. Table A2-2 enables us to answer the question of differences between the methods when summarized for all institutions. The average interview rating, based on 158 observations, is 3.1; the average role-playing rating is 2.9. The statistical tests of this difference indicate that it could well arise by chance alone, so that there is no evidence that students consistently receive higher ratings in one of the two methods.

The column on the right in Table A2-2, "Average Across Methods," shows some variation from school to school in the average rating, ranging from 2.5 to 3.5, with each average based upon a maximum of 60 observations. This variation is statistically significant, indicating that the schools do vary in the average rating they award. However, this variation across schools was essentially found for both methods in much the same way. Thus, the rank ordering of schools within interview is highly consistent with that within role-playing.

Question 5: How consistent are ratings in general from one method to another?

While the average level of rating is about the same for the two methods (interview vs. role-playing), the question of the consistency of ratings across methods is appropriate. Table A2-3 presents the number of

Table A2-2

Average Overall Ratings of Effectiveness as a Counselor for Students at Each of Nine Institutions

	Average Rating*				
Institution	Method A Interview	N	Method B Role Playing	N	Average Across Methods
A	2.8	15	2.4	15	2.6
B	3.6	14	2.8	15	3.2
C	3.6	8	3.3	10	3.4
D	3.3	12	3.2	15	3.3
E	3.5	15	3.5	15	3.5
F	3.1	13	3.2	13	3.1
G	3.2	16	3.1	14	3.1
H	2.9	14	2.8	14	2.9
I	2.9	16	2.8	16	2.9
J	2.4	15	2.6	15	2.5
Average Rating	3.1	138	2.9	140	3.0

*Each entry is the average across students of the overall rating of each student by two raters.

students at each of five levels for summed ratings derived from the interview and the role-playing. The potential range of these ratings is from 2, 2 ratings of "1" each, to 10, 2 ratings of "5" each. Thus, there was 1 student whose total rating on the interview was 2; this same student, who is indicated in the cell in the upper left-hand corner, received a summed rating of 2 on the role-playing, also. Similarly, the 3 students in the lower right cell received summed ratings of 9 or 10 for both the interview and the role-playing.

There is apparently a reasonable level of agreement between the two overall ratings. A correlation coefficient estimated from the basic data is .56. There are no students in the extreme cells at the bottom left or top right which would indicate a complete reversal of assessment. In a few instances shifts as great as three to five points are observed; for two students such shifts indicated a better role-playing appraisal, for four other students, a better interview appraisal. For all other students the two summed ratings were within three points of each other.

Question: 6 Are different aspects of interpersonal skill in counseling evaluated differently when different assessment methods are used?

Previous questions refer to the comparison of the overall ratings of counselor effectiveness. Question 6 focuses on the seven aspects of counseling skill which were considered in this study. Table A2-4 shows the average of the ratings of students by the two methods of assessment, interview and role-playing, for the seven traits. While the values cluster in a narrow range, the differences among these means are statistically significant, indicating that some traits are evaluated more generously than others. However, these patterns are highly similar for both interview and role-playing.

Table A2-3

Interpersonal Effectiveness in a Counseling Situation--Relationships Between Sums of Overall Ratings Derived From Two Methods of Assessment*

Interview Sum	Role-Playing Sum 2	3-4	5-6	7-8	9-10	Total
9-10	0	0	2	3	3	8
7-8	0	2	8	24	2	36
5-6	0	14	16	9	1	40
3-4	1	2	6	1	0	10
2	1	0	0	0	0	1
Total	2	18	32	37	6	95

[*] Number in cells represents number of students rated as indicated.

Table A2-4

Average Rating of Students in Analyses Samples Based on Two Methods of Assessment for Seven Different Counseling Skills

	Average Rating for All Students	
Type of Counseling Skill	Method A Interview	Method B Role Playing
1. Range of Counseling Experience	2.9	2.6
2. Ability to Deal with Unusual Problems	3.0	3.0
3. Knowledge of Counseling Methods	2.7	2.7
4. Effectiveness as Interviewer	3.1	3.0
5. Ability to Plan, Carry out Programs of Assistance	3.2	3.1
6. Communication Skills	3.3	3.1
7. Evaluative Skills	3.0	3.0

N's range from 87 to 94.

A somewhat different approach to the question of the impact of the assessment method on the ratings of the various traits is provided in Table A2-5, which reports the correlations between summed ratings on traits based on interview and summed ratings on traits based on role-playing. The number of cases available for each correlation varied from one measure to another. In general, the correlations indicate a degree of agreement that indicates that the two appraisals, while not totally consistent with each other, are broadly comparable. The levels of agreement do not differ greatly from one trait to another.

An important test of the usefulness of the seven traits is whether faculty are able to make consistent discriminations among students with respect to traits; that is, are certain students placed high on certain traits, others on different traits. Analysis 4 in Appendix 5b indicates that such distinctions are not made reliable. Thus, the seven traits may be helpful in directing judges' attention to relevant considerations of their task, but the seven ratings are best regarded simply as reflections of one overall impression of the students' interpersonal competence in the situation.

Question: To what extent do faculty within institutions agree
7 in assessing interpersonal skill in counseling?

The correlations reported in Table A2-5 are based on large groups of cases pooled across institutions. The variability of such correlations within institutions is indicated in Table A2-6, where the reliabilities of the two overall ratings--one based on interview, the other on role-playing--is reported, together with the correlation between them.

These coefficients are based on quite small numbers at eight institutions. In general, they indicate modest reliabilities and

Table A2-5

Correlations Between Summed Ratings Based on Interview and Role-Playing for Seven Traits for Available Data

Trait	Correlation	N
1. Range of Counseling Experience	.53	96
2. Ability to Deal with Unusual Problems	.57	96
3. Knowledge of Counseling Methods	.56	95
4. Effectiveness as Interviewer	.46	85
5. Ability to Plan, Carry out Problems of Assistance	.41	87
6. Communication Skills	.42	103
7. Evaluative Skills	.57	97

Table A2-6

Reliability Coefficients Showing the Degree of Faculty Agreement in Judging the Counseling Skill of Students -- by Two Methods of Assessment in Eight Institutions

	Reliability Coefficient					
Institution	Method A Interview	N	Method B Role-Playing	N	Correlation	N
A	.49	15	.82	15	.53	15
B	.82	14	.68	15	.67	14
D	.02	12	.50	15	.43	12
E	.62	15	.51	15	.69	15
F	.75	13	.74	13	.73	8
G	.85	16				
H	.92	14	.91	14	.84	13
I	.68	16	.69	16	-.16	16
Median R	.71		.69		.67	

consistency across traits. The level of reliability was much greater than that observed in the studies of A1, indicating that fewer problems were encountered with the observations than with the portfolios. The reliability of the interview at one institution seems atypically low; the correlation between methods at another is negative in spite of reasonable reliabilities. The results indicate that the appraisal of success at a given institution should be based on fairly substantial experience before acceptance or rejection, since results can vary so widely.

Question: 8 To what extent do individual students seem to do better in interview vs. role-playing simulations?

This question is an important one in the evaluation of interview and role-playing as methods of assessment. It focuses on issues which are central to the propriety and fairness of assessment procedure, probing the extent to which student performance might be affected by the method selected for appraisal. The analysis examined the rating given to 143 students at nine institutions. Essentially, it contrasts the variability among the differences in student ratings using the different methods of assessment with the general pattern of variability observed in rating, regardless of method.

There is no simple table which can present the variation among so many students for each method. Appendix 5, paragraph 7 presents the relevant analyses of variance, showing an interaction between student and method which is statistically significant. The Appendix also reports a value of .376 for the statistic Eta, which may be interpreted roughly like a correlation coefficient, and which suggests a tendency for some students to do better under one assessment method, and others to do better under the other. However, these data do not indicate whether students, if given the choice, would pick the method on which they do best.

While the general problem of limited sampling must be noted, the following observations may summarize the results:

1) The results of the appraisal of interpersonal skill by the two methods of interview and role-playing are quite reasonably consistent.
2) The reliability of each of the two methods, as appraised using small groups at a number of institutions, is quite reasonable.
3) The seven dimensions of interpersonal skill were not well differentiated. Each tends to reflect more the overall judgment of the rater than any special trait.
4) The rating of the individual student is significantly influenced by the method selected for appraisal.

Study A3 - WHAT IS THE EDUCATIONAL VALUE TO THE STUDENT OF THE ASSESSMENT OF INTERPERSONAL SKILLS?

The purpose of this study was to examine the effects of the assessment process on student personal growth and development with an emphasis on the educational goals of the student. In many areas, but perhaps uniquely in areas involving interpersonal skills, the assessment process itself has an integrative value which is reflected in students' reports of self-perceived levels of development. The precise way in which assessment helps learning is not well understood. It is plausible, of course, to suppose that the activity of assessment focuses and clarifies student perceptions and goals. This clarification makes it easier for students to evaluate themselves more accurately, to come to some sort of summary conclusions about themselves, and to find that in these insights lies the basis for a better future performance.

But this is essentially conjecture, and the present study had to be formulated without a precise control on the conditions which might govern variations in the amount of personal growth resulting from assessment. In consequence, while a positive finding of gain could support the general notion, a negative finding could not be seen as a conclusive demonstration that there is no educational value.

Description of Study

Each participating institution was asked to identify 20 students who were requested to complete questionnaires providing self-descriptions of status in terms of interpersonal competences: 10 of the students completing the questionnaire prior to assessment and 10 following it.

No overlap between the groups was permitted so that each group of 10 would be considered an independent sample from the respective populations of pre-assessment and post-assessment students.

The questionnaire posed five queries concerning interpersonal competence, asking the respondent to identify 1) major strengths in interpersonal skill, 2) major weaknesses in interpersonal skill, 3) how interpersonal skills relate to career goals, 4) how interpersonal skills related to goals for personal growth, and 5) how interpersonal skills relate to immediate educational planning. A copy of the questionnaire is presented as Appendix 4h. In devising these instructions, there was necessarily a tension between securing the fullest response from the student and yet holding the self-descriptions to a concise and manageable length. This lead to the use of somewhat paradoxical statements urging the student to respond "as fully but as concisely as possible," and to "try to confine your answer to the space provided," but "if you need more space attach additional sheets of paper." The typical student response to each question constitutes a brief paragraph.

Each questionnaire was evaluated by four judges. The four judges gave independent ratings to the student responses, using a three-point scale. A value of 12 constitutes a maximum, of 4, a minimum. Prior to these judgments, the raters were given a brief training exercise, rating model and answers secured in tryout samples. The judges were not affiliated with any of the institutions which supplied student data. All judges were experienced in assessing learning in nontraditional academic settings. Judges were, of course, unaware whether a given student belonged to the pre-assessment or post-assessment group.

Institutional Context

Data were incorporated from eight of the ten cooperating institutions. A number of these schools used the samples of A2, the students involved in the behavioral simulations, in the self-descriptive activity of A3. There were clearly a variety of differences in the student groups within different institutions. Thus, one institution reported all students to be over 25 years of age. In this, they probably constituted a unique group, different from the other institutions. The sample which was drawn from students at a women's correctional institution for Study A2 was used for Study A3 also. Again, this was an unusual group in many ways.

Many practical problems in fulfilling the design of the study were encountered. For example, the CAEL project technical staff had requested that students be alternately assigned to the pre- or post-assessment groups. But several institutions were required to modify this, since the assessment process, which often moved only at the will of the student, could stretch out beyond the CAEL-established request for a data report. Accordingly, these schools modified the sampling to insure conformity to schedule rather than to the quasi-experimental design.

The raters represented faculty at Thomas A. Edison College, Ramapo College, Temple University and Adelphi University. They worked independently, receiving the results from CAEL data, and from those institutions presenting student essays.

There were few reports of any difficulty on the part of the students in completing the questionnaire. The assessment process varied considerably from one institution to another. In some cases the assessment was actually for credit. In others, the granting of credit was not a factor except as a hypothetical possibility.

Data Analysis

A difficulty in the analysis is the distinction between the student responses and the rating of these responses. There can be no hope for determining the properties of the responses without adequate ratings. A first analytic task, then centered on evaluating the quality of the ratings.

Understandably, the rating process for this study proved to be quite difficult. The judgments of student responses were necessarily based on relatively brief evidence and a number of related assumptions. As in the use of portfolio evidence alone in Study A2, the absence of first-hand contact in Study A3 gave the judges considerable difficulty.

An illustration of the basic data for this study is presented in Table A3-1, which shows comparisons of ratings for students at a single institution. Note that observation is on a different student, and is associated with a unique student number, so that comparison within a row is not meaningful. These data, therefore, are most usefully compared by contrasting the average sum of ratings and the variability.

The level of rating at this institution shows no evidence for a gain or increase in effectiveness of self-description. The summed overall ratings for the post-assessment group are slightly greater, on the average, than those for the pre-assessment group, but the difference is not statistically significant. The basic analysis considered the average ratings awarded to pre-assessment subjects and post-assessment

Table A3-1

Summed Ratings of Pre-Assessment and

Post-Assessment Students at One Institution

Pre-Assessment		Post-Assessment	
Student Number	Rating Sum*	Student Number	Rating Sum*
1	8	11	4
2	4	12	4
3	6	13	4
4	7	14	8
5	5	15	5
6	8	16	7
7	5	17	5
8	5	18	5
9	4	19	4
10	6	20	5
Average	5.2		5.7

---

*Sum of four ratings on a three-point scale. Range of sum is 4-12.

subjects. The major analysis studied the variation in average rating at the participating institutions. A related analysis evaluated the individual questions which had together constituted the student response. In this related study, questions 1 and 2 were considered together by the judges.

Tentative Results

Question: 8 — Are students who have gone through an assessment of interpersonal skill more likely to describe their interpersonal competence in ways that indicate good self-awareness than is the case with students who have not participated in such assessment?

The procedure for answering this question is a two-step process. First, establish the meaningfulness of the ratings, then compare the average ratings awarded to pre-assessment and post-assessment students. The consistency of ratings is evaluated by a reliability coefficient. The sum of the four overall ratings (one for each judge) has a reliability of .82 when assessed by the analysis described in Appendix 5. Essentially, this analysis contrasts the extent to which the judges' ratings vary from one student to another with the extent to which they vary for the set of ratings given to a single student. If the contrast is not great, there is little evidence of reliability. For the ratings in this study, the judges gave very different ratings from one student to another, but much less often to the same student. Hence the high reliability. As hard as the judges reported it was to make the judgments, they did apparently define a concept of excellence in self-report that they could consistently agree on.

Do these consistent ratings vary for pre-assessment and post-assessment students? Table A3-2 presents an analysis for eight institutions showing the average sum of the overall rating both before and after assessment. In general the pattern does not support the idea that assessment influences ratings. The general effect is one of a close relationship between the pre-assessment level of rating and the post-assessment level within an institution. The difference between the overall average is not significant.

Student responses were elicited to five specific questions. The overall judgments were based upon the general impressions the judges received from these questions. The investigation of the extent to which the overall ratings show significant pre-assessment to post-assessment differences may be supplemented by a consideration of these individual questions. The judges were asked to rate each question on the three-point scale, except that the student responses to questions 1 and 2 were treated as a single response.

Table A3-3 shows the average judged ratings for each question for pre-assessment and post-assessment students and overall. It is evident that question 5 was somewhat different from the other questions, but that in no case was there successful differentiation between the pre-assessment and post-assessment students.

These contrasts are based on aggregate samples of 53 pre-assessment and 58 post-assessment students at six cooperating institutions. In the light of Tables A3-2 and A3-3, a more elaborate analysis, studying the usefulness of individual questions at individual institutions, did not seem justified.

This study attempted the ambitious task of seeking a differentiation between students who had not yet undergone assessment in interpersonal

Table A3-2

Average Summed Ratings for Pre-Assessment and Post-Assessment Students at Each of Eight Institutions

Institution	Pre-Assessment	N	Post-Assessment	N	Total	N
A	5.2	10	5.7	10	5.5	20
B	9.2	13	8.0	5	8.9	18
C	9.1	9	9.8	10	9.5	19
D	7.5	10	8.1	10	7.8	20
E	8.5	10	8.2	9	8.4	19
G	5.0	8	6.6	10	5.9	18
I	6.4	8	7.1	8	6.8	16
J	9.4	12	9.1	11	9.3	23
Overall	7.7		7.9		7.8	153

Table A3-3

Average Summed Ratings for Pre-Assessment and Post-Assessment Students at Six Institutions On Particular Questions and Overall

Questions	Pre-Assessment	Post-Assessment	Total
1 and 2 combined	7.7	8.2	8.0
3	8.2	7.9	8.1
4	7.8	8.2	8.0
5	7.1	7.0	7.1
Overall	7.7	7.8	7.8
N*	53	58	111

*These samples consisted of students at institutions A, B, C, D, G and I.

skills and students who had undergone such assessment. In spite of introductory training sessions, the raters reported considerable difficulty in making the appraisals. The evidence suggests, however, that the sum of the ratings is a quite reliable variable. When these ratings are used to contrast pre-assessment and post-assessment students, there is little evidence of any difference. Necessarily, given the distinction between student reports and the ratings they receive, it is possible that there are criteria for a successful differentiation. The study demonstrates, however, that ratings from four appropriate judges could not be used to differentiate in spite of an apparently reliable consensus on what was being rated.

Summary

1. To what extent can faculty differentiate various interpersonal competences in terms of their relevance to degree programs in different subject matter areas?

The general pattern of results in Study A1 did not demonstrate the ability of faculty to make useful differentiations in interpersonal competences. The raters in the field of psychology, at certain of the participating institutions, were most successful.

2. To what extent do different assessors agree on the level of interpersonal competence students have acquired and to what extent is there agreement between different methods of assessment?

Study A2 provides information bearing on these general questions and a number of related analyses. In general, assessor agreement was usefully demonstrated for both the interview and the role-playing simulations, and results secured by these two methods were reasonably consistent.

3. What is the educational value to the student of the assessment of interpersonal skills?

Study A3 in general failed to demonstrate evidence in support of interpersonal assessment itself as leading to growth in this area, in spite of the apparent success of the judges in assigning reliable ratings. The study used brief student self-descriptions as the basis for determining "educational value," a clearly limiting factor. Generalization, therefore, must be somewhat guarded. But the pattern of results obtained in this study was clearly negative.

Chapter V

STUDIES OF ASSESSMENT THROUGH THE PORTFOLIO

Warren W. Willingham and Joan E. Knapp

Nontraditional education places a great deal of emphasis upon promoting access to educational opportunity, recognizing learning wherever it occurs, and tailoring educational programs to the needs of students. These values argue for giving the student credit for what he or she has learned through experience prior to matriculation that may be directly relevant to a legitimate educational objective in the eyes of the institution.

Consequently, there has developed in recent years much interest in crediting prior learning, and several hundred institutions have such programs or are developing them. It is not unnatural that this sort of departure from traditional educational practice should raise a variety of questions concerning the adequacy of assessment procedures, their equity and educational soundness, protection of students and academic standards, and the role of institutions as credentialing agents. Observers want to know whether faculty have a clear idea of what is being credited, whether it is being assessed responsibly, and whether the whole process is defensible as a useful educational experience.

When the CAEL project was conceived in the fall of 1973, it was clear that the assessment of prior learning should be a major priority--both with respect to developmental work and validation studies that would begin to address such questions. In placing a priority on these objectives, the CAEL Steering Committee identified this work as "assessment of prior learning through the portfolio" because "portfolio" had come to represent a general technique for assembling information concerning prior learning, analyzing educational goals, and petitioning for credit.

Subsequently, CAEL stimulated various lines of developmental work in this area, including a handbook for faculty and administrators (Working Paper No. 6) and a student guide (Working Paper No. 7). The four studies reported in this chapter test assumptions and evaluate practices of the sorts described in those reports.

Students and faculty at the following 10 institutions participated in this field research:

Community College of Vermont
Empire State College
Governors State University
Metropolitan College, St. Louis University
New College, University of Alabama
San Francisco State University
Staten Island Community College
Texas Christian University
University of Missouri
University of Oregon

The names of individuals who coordinated the work at individual institutions are given on page v at the front of this report. Chapter III provides a general introduction to the field research, especially regarding its rationale, the general procedures followed, and suggestions regarding the interpretation of the findings.

Study B1 - DO FACULTY AGREE ON THE AMOUNT OF CREDIT AWARDED FOR PORTFOLIOS?

Portfolios presented by students seeking credit for prior learning vary a great deal. Frequently, if not typically, portfolios contain evidence of several types of learning that must be evaluated by different experts or different types of assessment in order to obtain an overall judgment as to how much credit should be awarded. Practices vary from institution to institution but to be fair and equitable, crediting prior learning must be consistent from student to student on the same campus. The primary purpose of this study was to determine the consistency of portfolio evaluation by determining the extent of agreement between two overall credit evaluations that have been arrived at independently. A second purpose was to examine the nature of institutional variation in credit awards.

Description of Study

This study involved two types of data from each institution: a) faculty evaluation of a set of local portfolios prepared by each institution's own students, and b) faculty evaluation of a set of illustrative portfolios that were common to the 10 participating institutions. These two types of data were especially relevant, respectively, to the primary and secondary purposes cited. Instructions and materials for the study are shown in Appendix 4-i through 4-1.

There were 14 illustrative portfolios containing evidence of learning in a variety of areas. At each institution each of these portfolios was submitted for evaluation to two independent sets of faculty experts, depending upon the nature of the learning involved. To facilitate the judging process, the relevant areas of learning were identified on each portfolio. For example, one illustrative portfolio contained learning outcomes in art and psychology. The two sets of faculty assessors, in this case, would each include an expert from psychology and from art who would independently evaluate the amount of credit due for learning in their respective fields. Thus, the two sets of faculty experts, working independently, produced two overall evaluations of the total credit awarded to each portfolio.* All 10 institutions provided data. They varied considerably in their experience in such assessment. One was completing its first 5 prior learning assessments this year; another now averages 1500 portfolio assessments a year. On most campuses judging the illustrative portfolios was necessarily different from the local assessment process since the student was not involved and the amount of information was less than that frequently if not typically available.

In addition to the analysis of illustrative portfolios, six institutions undertook a special duplicate evaluation of a set (typically 15) of their own local portfolios. Four institutions could not participate in this phase of the study because of various institutional restraints; e.g., one uses the portfolio process as a basis for developing contracts rather than awarding credit. Each of the remaining six institutions selected a representative set of portfolios of recent students and presented them to different faculty for a second, independent evaluation using local assessment practices. Again, the objective was to obtain two credit evaluations. That is, each was asked to carry out an independent but complete replication of their assessment process in order to obtain a second overall credit evaluation for each local portfolio.

---

*In a few instances, only one overall evaluation was obtained.

It is important to recognize the different research purposes for judging the local and the illustrative portfolios and the resulting implications regarding the procedures and materials used. In judging the local portfolios the importance of each institution using its own assessment procedures was emphasized because the main question concerned the extent of faculty agreement when the real operational assessment process is followed.

The illustrative portfolios served a different purpose. Some participants were initially concerned that the illustrative portfolios did not correspond to their own and that the judging did not include a transcript, the usual assessment process directly involving the student, and so on. But with the illustrative portfolios the question is not how good the portfolios were or whether the assessment process was complete (as far as we can tell, they were neither exemplary nor atypical of portfolios generally; Appendix 4j shows an example). The question of interest regarding the illustrative portfolios is rather how much do institutions vary in their evaluation of the same learning described in the same way? The consistency of faculty judgment of illustrative portfolios within institutions was examined as well in order to be sure that variations across institutions were not simply a function of poor judgment of poor materials. (As we shall see, the illustrative portfolios were judged as consistently within institutions, as were the local portfolios.)

Data Analysis

The local and illustrative portfolios were analyzed separately. In both analyses the dependent variable was total credit awarded the student. Credit was translated to semester hours. The data were analyzed across institutions to determine whether there were significant differences in the amount of credit awarded (generally and with respect to particular portfolios). Data were analyzed within institutions to determine how consistent the overall credit evaluations were (see Appendix 5d).

As an example of the basic data of the study, Table B1-1 shows evaluations for illustrative portfolios at one institution. The table indicates the result of two independent evaluations of each portfolio. It should be remembered that different departments and different faculty were involved in evaluating the various portfolios. We are not concerned here with which disciplines were involved or whether members of the same discipline agreed, but rather whether the overall assessment process yields two independent total credit evaluations that are in reasonable agreement. In the sample data of Table B1-1 there is relatively little overall discrepancy between the first and second evaluation on the average, but the two evaluations do yield different results in the case of individual portfolios. The reliability coefficient of .73 describes the extent of agreement; the overall average credit award of 11.4 represents the faculty judgment as to how much credit these portfolios might earn at this institution assuming there were no restraints on the total amount of credit allowable. These are the sort of summary statistics that speak to the questions addressed in the following paragraphs.

Table B1-1

Sample Data Showing Total Credits Awarded in Two Independent Evaluations of 14 Illustrative Portfolios in One Institution

	Total Credit Awarded Each Portfolio		
Illustrative Portfolio	First Evaluation	Second Evaluation	Average for Two Evaluations
1	19	10	14.5
2	13	6	9.5
3	0	3	1.5
4	29	12	20.5
5	17	24	20.5
6	9	0	4.5
7	21	30	25.5
8	9	0	4.5
9	0	0	0
10	3	2	2.5
11	13	16	14.5
12	15	12	13.5
13	15	8	11.5
14	13	20	16.5
Average	12.6	10.2	11.4
Reliability			.73

Results

Question: 1 To what extent do faculty within an institution agree on credits awarded on the basis of a portfolio assessment process?

The degree of faculty agreement in awarding credit to portfolios is best represented by the reliability coefficients shown in Table B1-2 (Analysis 2, Appendix 5d). In the case of both the illustrative portfolios and the local portfolios, the median coefficient lies in the low to mid-seventies. In a couple of instances the coefficients are unusually high or low, but most lie well within the range of expected fluctuation (which is necessarily wide due to the fact that the samples are small). The one very high coefficient possibly reflects a lack of complete independence between the first and second ratings at Institution B; the single zero coefficient may well be atypical for Institution E. The median values cited are not much affected by such extreme and likely unrepresentative cases. The general level of these reliabilities, as indicated by the medians, is considerably below that of well-developed objective examinations, but does not appear much different from reasonably careful expert judgment as exercised in similar situations (see "Interpreting the Results" in Chapter III).

Question: 2 To what extent do institutions vary in the amount of credit they award for prior learning?

Table B1-2 shows average credit awards for both illustrative (those common across institutions) and local portfolios. While there is no assurance that the sets of local portfolios are representative of the 6 institutions shown, it is evident that the average amount of credit granted varies quite widely (from 5 to 54 hours). Since the "local" portfolios were assessed with local procedures, there may be a number of

Table B1-2

Average Credits Awarded and Reliability of Faculty Judgments for Illustrative (Common Across Institutions) and Local Institutional Portfolios

	Illustrative Portfolios			Local Portfolios		
Institution[#]	N*	Average Credit	Reliability	N*	Average Credit	Reliability
A	14	40.2	.87			
B	14	24.8	.79	20	26.8	.99
C	14	11.4	.73	15	4.8	.88
D	9	25.0	.73	15	6.1	.67
E	12	16.2	.00	15	54.2	.59
F	13	9.5	.55			
G	9	18.3	.71	15	5.9	.85
H	14	8.2	.69			
I	14	6.9	.52			
J	14	24.1	.80	13	18.2	.53
Median		17.2	.72		12.2	.76

[*] Number of portfolios.

[#] Institutions A-D judged a slightly different set of illustrative portfolios than did Institutions E-J (see note on Table B1-3).

reasons for these variations in average credit awards: institutions impose different limits on prior credit; different institutions attract different types of students; the portfolio evaluation process may include forms of credit at one institution (e.g., transfer credit) not included at another; and so on. In any event some colleges clearly grant very limited amounts of credit for prior experiential learning.

Since the illustrative portfolios were common across institutions and judges were instructed to assume no limit on the amount of credit that could be awarded, these common portfolios offer a more controlled test of the extent of institutional variability in evaluating prior learning.

Table B1-2 indicates that there are substantial differences in the extent to which different faculties are willing to grant credit for the same portfolio describing the same learning. The credit awarded the set of illustrative portfolios differed significantly among institutions from a statistical as well as a practical standpoint (Analysis 1, Appendix 5d). Some faculties consistently judged the portfolios to be worth four to five times as much credit as did other faculties. Presumably, these differences reflect an institutional philosophy and conscious policy concerning the sorts of experiential learning represented in the set of common portfolios which merit credit.

Question: 3 Do different institutions evaluate the same portfolio differently? (aside from the general tendency of some institutions to grant more credit than others)

By implication, this question asks whether different philosophical orientations of institutions are reflected in a tendency to grant more credit to some types of portfolios than to others. Relevant data appear in Table B1-3, which shows average credit evaluations for each of the 14

Table B1-3

Average Credits Awarded for Illustrative Portfolios by Each of Six Institutions*

	Illustrative Portfolio														
Institution	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Average
A	71	30	1	68	73	57	67	6	3	7	25	24	43	84	40.2
B	40	24	3	33	51	30	22	9	2	16	52	15	16	30	24.8
C	14	9	1	20	20	4	25	4	0	2	14	13	11	16	11.4
H	12	7	0	0	22	9	13	7	1	4	10	7	1	12	8.2
I	7	3	0	15	12	2	15	3	0	4	3	0	12	18	6.9
J	36	15	0	30	46	10	28	12	10	46	26	12	22	39	24.1
Average	30.3	15.0	1.1	27.9	37.6	18.9	28.7	7.0	2.7	13.6	22.0	12.0	17.6	33.4	

*Institutions A, B, and C evaluated a slightly different set of portfolios than did Institutions H, I, and J. Both sets of 14 portfolios involved the same documentation of learning but the two sets differed somewhat in certain "extraneous" student characteristics (e.g., age) unrelated to this analysis. (Averages do not agree exactly with detail because decimals were dropped.)

illustrative portfolios by each of 6 institutions that provided complete data. It is evident that some portfolios were consistently awarded substantially more credit than others, and as noted in the previous table, some institutions tend to make higher credit awards (see Analysis 1, Appendix 5d). Even taking those two facts into account, there was still a significant tendency for institutions to view different portfolios differently. For example, Institution I rated Portfolios 5 and 10 the same (both awarded 46 credits), but Institution H evaluated those two portfolios quite differently (22 versus 4 credits)--presumably because of the different content of the two portfolios. But the example illustrates the much bigger source of variation--the fact that some institutions made substantially higher evaluations across the board than did others.

In the foregoing discussion two types of discrepancies were identified. One was inconsistency among raters within a single institution. This is the "error" a student might expect to result from one judge rather than another evaluating his or her portfolio. Note that the same potential error applies to a team rating since, in theory, it is always possible to empanel a different team whose award would not likely agree exactly with that of the first team. The second type of discrepancy is the difference that can be expected among awards of different institutions. This is the "error" that a student can expect if he or she submits a portfolio to a judge at one college rather than a judge at another college.

Table B1-4 summarizes both types of discrepancy (see Analysis 4, Appendix 5d). It is apparent that judgments are substantially more at variance when compared across pairs of judges at different institutions than for pairs of judges within the same institution. In the former case

Table B1-4

Frequency Distribution of Absolute Differences Between Two Credit Awards for the Same Portfolio

	Proportion of all differences between two awards	
Differences Between Two Awards	Between Raters Within the Same College	Between Raters in Different Colleges
0	13.9	8.9
1-5	31.6	19.6
6-10	27.8	20.9
11-20	15.2	17.4
21-30	7.6	11.7
31-40	1.3	7.9
41-50	1.3	5.7
Over 50	1.3	7.9
Total Comparisons	79	316
Median Difference	6.3	12.2

the median difference between judgments is 12.2 credit hours; in the latter case the median difference is 6.3. Furthermore the really large discrepancies are mostly between institutions. Thus the data support the assumption that students may receive substantially different credit awards, depending upon the choice of institution to which a portfolio is presented.

Study B2 - DO FACULTY AGREE ON WHAT CONSTITUTES COLLEGE-LEVEL LEARNING?

When students submit various types of prior experiential learning for credit, some judgment must be made as to whether the institution considers specific instances of learning to be college level or not. Institutions will hold different views, but a basic requirement for the validity of crediting practices is that faculty within an institution agree reasonably well as to which types of learning are creditable at that particular institution. The primary purpose of this study was to examine the extent to which faculty do agree on what types of prior learning and competencies are creditable. A secondary purpose was to examine the nature of institutional differences in judgments regarding what constitutes college-level learning.

Description of Study

Instructions and materials for the study are shown in Appendix 4i and 4m-n. Five faculty members from each of ten participating colleges were asked to judge independently each of fifteen instances of learning. Each instance included a brief description of the type of experience from which the learning was acquired and an indication of the type of evidence of learning that was available. Raters were not asked to assign credits or to assess the specific instance of learning but simply to indicate whether they considered each type of learning creditable and, if so, at what educational level.

There are two important qualifications to be expressed about the design of this study. One is the limited amount of information available to the judge (Table B2-3 identifies the 15 items; Appendix 4m shows what information the judges actually received). Another is the fact the same five faculty judged all 15 items rather than each item being referred to an expert in that particular type of learning outcome.

Undoubtedly, both of these considerations limit the agreement one can expect to find, but it is important to recognize that we are not seeking here expert assessment but simply whether there is reasonable agreement within a faculty as to whether different types of learning ought to be considered creditable. The real question then is whether a college faculty in general should be in reasonable agreement on the creditability of such learning outcomes. The authors would maintain that the answer is yes since the learning instances selected vary widely in depth of learning, but the reader must judge the validity of that assumption.

Data Analysis

The dependent variable used in the analysis was a four-point rating: 0 - not creditable at college level; 1 - lower division credit; 2 - upper division credit; 3 - graduate level credit. The data were analyzed within institutions to determine how well faculty agree on what is creditable. The data were analyzed across institutions to determine whether different faculties reflect different institutional orientations regarding what is creditable (see Appendix 5e). Virtually complete data were available for this study from all ten participating institutions.

As an example of the data, Table B2-1 shows one set of faculty ratings from one institution. It puts two sources of rater disagreement in perspective. One is difference among raters. For instance, Rater II shows a systematic tendency to rate learning outcomes as worthy of upper division or graduate credit (11 of 15); on the other hand, Rater I judged only 3 of the same 15 examples as being worthy of credit at that level. Another source of error is inconsistency; i.e., the fact that different

Table B2-1

Sample Data From One Institution Showing Five Faculty Ratings as to Whether Specific Instances of Learning Deserve College Credit (0 - Not Creditable; 1 - Lower Division Credit; 2 - Upper Division Credit; 3 - Graduate Credit)

	Faculty Ratings					
Learning Instance	I	II	III	IV	V	Average Rating
1	0	2	2	1	1	1.2
2	2	3	2	2	1	2.0
3	1	3	1	1	1	1.4
4	0	2	1	1	1	1.0
5	2	3	3	3	2	2.6
6	1	1	1	1	1	1.0
7	1	1	1	2	2	1.4
8	1	2	1	1	1	1.2
9	1	1	1	2	1	1.2
10	2	3	2	2	2	2.2
11	1	2	1	1	1	1.2
12	0	2	2	2	2	1.6
13	0	3	0	1	1	1.0
14	1	1	1	1	1	1.0
15	1	2	2	2	1	1.6
Average	.9	2.1	1.4	1.5	1.3	1.4
*Interrater						
Agreement: R_a						.41
Consistency: R_b						.56

*Estimates are for two rater composite. R_a includes both inconsistency in ratings and consistent differences among raters as error. R_b does not include the latter as error.

raters do not rank order the 15 instances of learning similarly with respect to their credit worthiness. At the bottom of Table B2-1 two estimates of inter-judge agreement are shown. One correlation (R_a) includes both of these types of disagreement as error. The other (R_b) includes only inconsistency as error and ignores the consistent difference between raters. The last column of Table B2-1 shows the summary data for this institution. That last column in Table B2-1 appears as Institution B in Table B2-2, which reports similar data for the other nine institutions as well.

Results

Question: 4 To what extent do faculty within institutions agree in identifying which specific instances of learning are worthy of college credit?

Table B2-3 shows the 15 learning instances and indicates the average rating for each. It is evident that there is at least some agreement among faculty that these learning instances are differentially creditable because the ratings range from well below lower division credit (.5) to a level approaching graduate credit (2.5). But how well do faculty within an institution agree in their judgments over the full set of 15 instances?

The last two rows of Table B2-2 indicate the overall degree of agreement among the faculty at each institution. R_a includes both types of error described above and the median coefficent is .53 (see Analysis 1, Appendix 5e). A coefficient of this size would indicate, for most purposes, only modest agreement. The reader can judge that more directly by referring again to Table B2-1 where the sample data illustrate fairly well this extent of agreement (and disagreement). R_b does not include the "rater-error;" i.e., it assumes that one can control such error fairly easily. In this

Table B2-2

Average Rating of the Level at Which Specific Instances of Learning are Judged Creditable and the Degree of Interrater Agreement in Those Ratings (Rating of 15 Instances at 10 Institutions for Which: 0 - No Credit; 1 - Lower Division Credit; 2 - Upper Division Credit; 3 - Graduate Credit)

Learning Instance		Institutions									
		A	B	C	D	E	F	G	H	I	J
1		.4	1.2	.4	.8	.6	.6	.2	.4	.0	.8
2		2.2	2.0	1.8	2.2	1.8	2.4	1.8	2.2	1.8	2.2
3		1.6	1.4	1.4	1.2	.8	2.0	1.4	1.3	1.2	1.4
4		1.0	1.0	1.0	.8	.8	1.4	1.0	.6	.8	1.2
5		2.6	2.6	2.4	2.4	2.4	2.6	2.4	2.8	1.8	2.6
6		1.6	1.0	1.4	1.4	1.4	1.4	1.2	1.4	1.6	1.4
7		1.4	1.4	1.2	1.8	1.6	1.4	1.6	1.8	1.0	1.8
8		.8	1.2	.8	1.0	.8	.8	1.0	1.2	1.0	.8
9		1.2	1.2	1.8	1.4	1.0	1.0	.8	2.0	.6	1.8
10		1.4	2.2	1.8	2.6	1.8	1.4	1.8	2.2	.4	2.6
11		.8	1.2	1.4	1.2	1.0	1.0	1.0	1.6	1.0	1.0
12		1.4	1.6	2.0	2.4	1.6	2.4	2.0	2.6	.3	2.2
13		1.0	1.0	.8	1.0	.2	.8	.4	1.8	.2	.8
14		.8	1.0	1.2	1.0	1.0	1.2	1.4	2.0	1.0	1.2
15		.8	1.6	1.4	1.2	1.2	.8	.8	1.6	.0	1.2
Average		1.3	1.4	1.4	1.5	1.2	1.4	1.3	1.7	.8	1.5
*Interrater Agreement:	R_a	.45	.41	.18	.53	.64	.43	.54	.58	.64	.56
Consistency:	R_b	.50	.56	.33	.68	.74	.48	.68	.63	.65	.62

*Estimates are for two rater composite. R_a includes both inconsistency in ratings and differences in leniency among raters as error. R_b does not include the latter as error.

Table B2-3

Fifteen Learning Instances and the Average Rating Regarding the Creditability of Each (0 = No Credit; 1 = Lower Division Credit; 2 = Upper Division Credit; 3 = Graduate Credit)

Learning Instance	Average Rating
1. Knows the roles and functions of the local YMCA.	.5
2. Knows and can apply the principles of cost accounting.	2.0
3. Knows and can apply techniques of written business communication.	1.4
4. Knows and can apply principles in meal planning and serving.	1.0
5. Knows and can apply techniques of fiction writing.	2.5
6. Knows German, can communicate with native speakers.	1.4
7. Knows causes of abnormal personality and therapeutic methods.	1.5
8. Competent in secretarial skills and office management.	.9
9. Can apply principles of group dynamics in a group setting.	1.3
10. Can apply EEO policies in a major affirmative action program.	1.8
11. Understands basic tenets of major religious groups.	1.1
12. Can plan and conduct recreational and therapeutic activities for all age groups.	1.9
13. Knows the function and application of standard metal fabricating machines.	.8
14. Knows and can apply basic methods of piano and accordion playing.	1.2
15. Can apply management and administrative skills in a church setting.	1.1
Total	1.4

case the median goes up to .63--in the predictable direction but not a large change. It would seem that the main source of disagreement is difference of opinion about individual types of learning, not a marked tendency for some raters to judge consistently higher or lower than other raters.

Question: 5 To what extent do institutions vary in their general tendency to regard learning as college level?

The "Average" line of Table B2-2 shows the average rating for each institution over all judges and all instances of learning. These average ratings vary from .8 to 1.7. This means approximately that the highest average of these 10 institutions rates these instances of learning as creditable on the average at almost 2 years beyond the level represented by the lowest overall average rating (e.g., upper division vs. lower division). These differences would appear to be educationally significant; there is no question of their statistical significance (Analysis 2, Appendix 5e).

Question: 6 Do different institutions evaluate the creditability of specific types of learning differently (aside from the general tendency of some to rate higher than others)?

It is evident in Table B2-2 that there is a good deal of variation in the way individual institutions judge the 15 learning instances. It is also clear that some institutions rate higher than others. If this latter variation is taken into account (the column averages), there remains no statistically significant relationship between institution and learning instance (Analysis 2, Appendix 5e). Thus, the answer to the question is no. These data indicate that institutions vary largely in the general tendency to rate high or low in judging what is creditable and do not bring consistently different judgments to bear on different types of learning (at least with respect to the types of learning cited in the instances used in this study).

Study B3 - TO WHAT EXTENT DO FACULTY AGREE ON THE TYPE OF DOCUMENTATION REQUIRED AS EVIDENCE OF EXPERIENTIAL LEARNING, AND TO WHAT EXTENT DO VARIOUS TYPES OF DOCUMENTATION PROVE AUTHENTIC?

A student claim to prior learning, depending upon its nature, may be assessed directly and/or substantiated with documentation. There are many possible types of documentation, and documenting often serves somewhat different purposes depending upon the nature of the learning. What constitutes adequate documentation is often a matter of judgment. This study examines the degree of consistency among faculty with respect to the nature of documentation they judge to be required for various examples of experiential learning.

Description of Study

Research materials and instructions for this study are shown in Appendices 4i, 4o, 4p, and 4q. Fifteen examples of learning outcomes were excerpted from student portfolios. Each was described in about 100 words. The set of 15 learning outcomes were considered independently by 5 experienced faculty members from each of 10 participating institutions. For each of the examples of prior learning, faculty were asked to check which of the following types of documentation they would normally require for assessment.

- Letter
    - --Letter of verification that experience took place (from supervisor, client, or co-worker)
    - --Letter of evaluation of performance (from supervisor, client, or co-worker)
    - --Letter of commendation
- Work Description
    - --Official job description
    - --Student explanation of task performed
    - --Course outlines, syllabi, and/or assignments
- Professional Recognition
    - --License
    - --Certificate
    - --Membership in professional organization
- Work Product or Record
    - --Products or work samples such as plans, tapes of performance, manuscripts, memos, reports, etc.
    - --List of books read

In a second phase of this study, institutions were also asked to provide verification data on documentation found in a sample of 25 local student portfolios. All 10 institutions participated in the first phase; 8 submitted data for the second phase.

Data Analysis

The various types of documentation listed above were grouped into four categories as indicated: Letter, Work Description, Professional Recognition, and Work Product or Record. If a faculty judge checked any of the three items in the "Work Description" category, his rating was recorded as "yes" for that category and the particular type of learning in question. Thus, the basic data of this study were yes/no judgments by each rater for each of the 4 types of documentation for each of the 15 examples of learning. For purposes of analysis, the yes/no judgments were coded as one/zero. The analysis is described in Appendix 5f.

A sample of one institution's data for the "Letters" category is shown in Table B3-1. In this particular sample of data, the raters do not differ greatly in their overall tendency to require letters as documentation, but there are numerous instances of disagreement regarding individual learning examples. It is within such sets of ratings that the data were analyzed to estimate the extent of faculty agreement regarding documentation requirements. The pooled judgments of faculty (the last column of Table B3-1) were analyzed in various ways to examine differences across institutions and types of learning.

Table B3-1

Sample Data From One Institution Showing Whether Each of Five Faculty Raters Judged Some Type of Letter to be Required Documentation for Each of 15 Examples of Prior Learning*

	Faculty Raters					
Learning Example	A	B	C	D	E	Proportion Yes
1	y	y	y	y	n	.80
2	y	y	y	y	y	1.00
3	n	n	n	n	y	.20
4	y	y	y	y	y	1.00
5	n	y	n	y	y	.60
6	n	n	n	y	n	.20
7	n	y	y	n	y	.60
8	y	y	y	y	y	1.00
9	y	y	y	y	n	.80
10	y	y	y	y	y	1.00
11	y	y	y	n	y	.80
12	y	y	y	y	y	1.00
13	n	y	n	y	n	.40
14	y	y	y	y	y	1.00
15	y	y	y	y	y	1.00
Proportion Yes	.67	.87	.73	.80	.73	.76

[*] y = yes; n = no. Reliability of interjudgment agreement for these data was .49 (for a 2-rater composite).

Results

Question: 7 To what extent are different types of documentation considered to be required for assessment of specific types of prior learning?

Table B3-2 shows the proportion of all faculty raters who indicated that each of the four types of documentation were required for each of the fifteen examples of learning. As would be expected, letters were the form of documentation most often required, though work products and records were not far behind. Formal professional recognition such as a license was considered a required form of documentation least frequently. Inspection of Table B3-2 indicates a clear tendency for the preferred form of documentation to depend upon the character of the learning (see Analysis 2 in Appendix 5f). In each of the 15 examples at least 1, and usually 2, forms of documentation are indicated by the great majority of raters. The pattern of preferences seems, in most instances, to make sense, though some results can easily be questioned (e.g., why 30% of the raters felt that a letter should be required documentation for the competence represented in having published two books).

Question: 8 Are there systematic differences among institutions in the extent to which they require different types of documentation?

Table B3-3 indicates that these institutions are all rather similar in the extent to which they require students to provide documentation. Only Institutions A and I show a generally lower tendency to require documentation of the sorts listed. There was a statistically significant tendency for different institutions to prefer different forms of documentation, but such differences were not large (See Analysis 2 of Appendix 5f). The extent of institutional variation in requiring documentation was roughly comparable for each of the four types shown in the four columns of Table B3-3 (see Analysis 3 of Appendix 5f).

Table B3-2

Proportion of All Faculty Raters (N=50) Who Indicated That Each of Four Types of Documentation Was Required for Each of 15 Examples of Prior Learning

Examples of Learning	Letter	Work Description	Professional Recognition	Work Produ Or Record
1. Reading music and playing musical instruments; pitch and sight reading.	.90	.24	.26	.80
2. Ability to conduct evaluation interviews as a result of a course.	.86	.88	.34	.44
3. Basic understanding of the American West, cowboy folklore, American Indian, as a result of reading.	.46	.36	.02	.94
4. Knowledge of negotiating labor union contracts.	1.00	.68	.22	.50
5. Know and can apply basic theories, techniques of swimming and water safety as a result of a course.	.58	.40	.78	.66
6. Understanding of ethical philosophy as a result of writing two books.	.30	.32	.06	.94
7. Development of skills needed in opinion research as a result of self employment.	.78	.54	.12	.92
8. Ability to teach principles of science and philosophy in a course in respiratory theory.	.94	.78	.34	.68
9. Understanding the theory and philosophy of Jail Ministry.	.84	.62	.22	.78
10. Ability to assess the geriatric patient and provide care as indicated.	.96	.70	.48	.44
11. Acquainted with techniques of pastoral counseling particularly pre-marital counseling.	.86	.56	.20	.70
12. Knowledge and application of computer software in a large company.	.94	.74	.16	.70
13. Development of job description-classification system for salary administration for a small plant.	.46	.44	.06	.80
14. Knowledge of criminology and sociology as a result of experience as a police officer.	.94	.76	.20	.44
15. Knowledge of parent-effectiveness and application to student unrest situations.	.90	.66	.14	.70
Overall Proportion	.78	.58	.24	.70

Table B3-3

Within Each Institution, the Relative Frequency* With Which Each of Four Types of Documentation Was Required

	Category of Documentation			
Institution	Letters	Work Descriptions	Professional Recognition	Work Product Or Record
A	.69	.39	.17	.61
B	.79	.64	.17	.68
C	.88	.64	.23	.76
D	.79	.61	.28	.68
E	.84	.73	.27	.71
F	.84	.56	.39	.67
G	.76	.49	.28	.61
H	.84	.53	.28	.84
I	.63	.48	.08	.63
J	.76	.71	.25	.77
Total	.78	.58	.24	.70

*Each entry represents the proportion of "yes" judgments for all raters and examples.

Question: To what extent do faculty agree on what documentation
9 is required for specific types of learning?

Table B3-4 shows how well faculty within individual institutions agree about documentation requirements. On the whole these coefficients indicate quite poor or, at best, a modest degree of agreement among faculty within each institution. The low coefficients suggest that faculty quite often do not agree as to whether a particular type of documentation is appropriate for a particular instance of learning. Judges were more likely to agree with respect to "Letters," but even there the agreement is not high. The wide variation in the coefficients suggests that some institutions may have given attention to documentation guidelines in some areas but not others.

Question: To what extent are various types of documentation
10 found to be authentic?

Even though the participating institutions made a conscientious effort to carry out this aspect of the study according to instructions, it does not appear that the data are useful in answering the question posed.

Most of the initial data showed almost 100% authenticity. Much of the data subsequently received were ambiguous or incomplete. Two sets reflected a low rate of authenticity. It appeared from the data and from comments of site coordinators that participants may have had quite divergent views of what constituted authenticity.

Institutions ran into various other difficulties in obtaining the necessary information: 1) some of the documentation was relatively old and sources were difficult to locate; 2) some students did not want former employers and colleagues who wrote letters to be contacted; 3) some students were troubled at the thought that their integrity was being questioned, so the institution withdrew from the verification procedure; and 4) in at least one institution documentation is not retained but returned to students during assessment.

Table B3-4

The Extent of Faculty Agreement as to Whether Different Types of Documentation Should be Required for Assessment*

	Reliability Estimate for:			
Institution	Letters	Work Descriptions	Professional Recognition	Work Product Or Record
A	.53	.00	.00	.21
B	.37	.12	.37	.12
C	.42	.06	.46	.00
D	.12	.30	.53	.32
E	.30	.00	.00	.00
F	.55	.18	.16	.30
G	.12	.11	.37	.21
H	.60	.00	.11	.00
I	.09	.20	.00	.57
J	.49	.21	.56	.12
Median R	.40	.12	.26	.17

*Entries are reliability estimates for a composite based upon two ratings.

Despite all the work devoted to gathering the data, it did not seem defensible to analyze the data for the reasons cited. A rigorous view would lead one to conclude that such post-facto checks on authenticity are probably impractical and unconvincing since any instances of improper documentation might remain hidden in various exceptions that prove impossible to follow up later. It seems likely that a different type of study is required to obtain useful information concerning authenticity of documentation.

Study B4 - WHAT IS THE EDUCATIVE VALUE TO THE STUDENT IN DEVELOPING A PORTFOLIO AND HAVING IT ASSESSED?

Assessment of experiential learning is not conducted solely for the purpose of awarding credit to students. A basic principle of the CAEL project is that the assessment process itself should have direct educative benefit to the student. Assessment should help the individual understand her or his competence and how that competence is related to personal objectives and plans; i.e., assessment should lead to improved self-awareness. The purpose of this study was to determine whether portfolio assessment has such an effect. The study is based upon the following rationale. If portfolio assessment does have direct educative benefit, then we assume that benefit can be discerned in the self-awareness students exhibit in their free responses to pertinent questions. That assumption was tested as follows.

Description of Study

Eight participating institutions selected representative but separate samples of students who had recently gone through the local portfolio assessment process (post-assessment group) or were planning to do so (pre-assessment group). For some institutions with relatively small programs, the sample consisted of most such students available in the data collection period, January to March 1976. In most instances participating institutions reported difficulty in obtaining student cooperation because they were hard to contact or resisted completing "more documents." Some site coordinators felt that the post-assessment group took the task less seriously and this may have biased the data against finding the effect hypothesized. Students were asked to respond in writing to the following questions:

1. If you have had any previous experiences in work, community or church activities, hobbies, and/or travel which you feel resulted in learning that is worthy of college credit, please list them and what you learned from them below.

Experience	Learning Outcome

2. For each of the experiences listed above explain how you could demonstrate to a faculty member of this college that you indeed learned what you said you learned.

3. Briefly explain how your life experience learning is related to your career goals.

4. Briefly explain how your life experience learning is related to your educational goals.

5. Briefly explain how your life experience learning is related to your personal goals.

The students' responses were returned to ETS identified as to pre- and post-assessment group. They were rated as described below. In addition, institutions completed a brief form describing distinctive aspects of their assessment program (Appendix 4s), but this information was not analyzed systematically because the ratings of student responses did not suggest that there were any consistent differences among institutions in the educative effects of assessment.

Data Analysis

Upon arrival, all student questionnaires were coded and all identifying information was removed. They were then arranged in random order to further remove any extraneous clues as to who the student was, where he or she had come from, or whether the individual had or had not gone through a portfolio assessment. The students' responses were rated by four faculty experts who represented different experiential learning programs and were conversant with portfolio assessment (though none had participated in the study)*. The four raters met at ETS and, prior to any rating, discussed what constituted

---

*The judges were: Martha D. Sachs, Ramapo College
Irene G. Casper, Temple University
Keitha Donnelly, Adelphi University
Paul I. Jacobs, Thomas A. Edison College

good and poor responses. Responses to each question were judged according to the following criterion:

> To what extent does this response reflect the self-awareness that one would expect of a person who has a mature understanding of personal competence and goals?

The response of each student to each question was rated on the following three-point scale:

3 - Highly developed self-awareness

2 - Moderately developed self-awareness

1 - Little evidence of self-awareness

Raters also made one overall rating of each student's five responses taken together. These several ratings constituted the data of the study. Pre-assessment and post-assessment groups were compared both within and across institutions to determine whether such "blind" ratings of student responses would suggest a significant difference in self-awareness between the two groups.

Results

Question: 11 — Are students who have been through the portfolio assessment process better able to describe their learning and how it relates to their goals than are students who have not been through the process?

Table B4-1 shows average overall ratings for pre- and post-assessment groups at each of the participating institutions. Some of the groups are quite small, but in most institutions the post-assessment group received a higher rating. For the total group the difference (1.85 versus 2.14) is statistically quite significant (Analysis 3, Appendix 5g). The practical significance of this difference is somewhat easier to interpret if one compares the percentage of high ratings in each group. Whereas 22% of the

Table B4-1

Average Rating of Papers for Independent Groups of Students at Each Institution Who Had and Had Not Participated in Portfolio Assessment

	Pre-Assessment Group			Post-Assessment Group		
Institution	N	Average Rating	% Rated High*	N	Average Rating	% Rated High*
B	20	1.9	20%	11	2.1	36%
C	5	2.1	20	4	2.1	25
D	10	1.9	20	7	2.2	36
F	27	1.8	16	14	1.9	19
G	13	2.1	40	7	2.8	79
H	12	1.8	19	9	1.9	30
I	8	1.4	6	3	2.7	75
J	9	2.2	39	6	2.2	29
Total	104	1.85	22%	61	2.14	37%

---

*Student papers were rated on a 3-point scale where "3" was high. These data are based upon the "overall" rating of each of four raters. "% Rated High" refers to the percentage of "3" ratings in each group.

students in the pre-assessment group were rated high, the corresponding number was 37% in the group that had been through portfolio assessment.

Question: 12 Is the apparent "positive effect" of assessment greater at some institutions than at others?

There were differences across institutions if one looks at the overall ratings for the total sample from each institution. The responses of students from some institutions were rated generally higher than those from other institutions. But there was no statistically significant tendency for the difference between pre- and post-assessment groups to be greater at one institution than another (Analysis 3, Appendix 5g). The fact that the pre/post difference appears larger in some cases than in others can be assumed to represent chance fluctuations due to the small size of the groups.

Question: 13 Do responses to specific questions suggest that students who have participated in portfolio assessment have gained more in certain aspects of self-awareness than in others?

Judges were able to make quite reliable distinctions between the five questions (see Analysis 2, Appendix 5g). Also, Table B4-2 indicates that, overall, there was a greater percentage of "high" ratings in the post-assessment group for some questions than others (Analysis 4, Appendix 5g), but this result may merely reflect a tendency for raters to give both pre- and post-assessment students higher ratings on some questions. The table also indicates that for every question, there was a greater percentage of "high" ratings in the post- than in the pre-assessment group. The pattern of that difference did not vary significantly, however, from question to question (see Analysis 5, Appendix 5g). That is, one can assume that the observed advantage of the post-assessment group is comparable for the five questions.

Table B4-2

Percentage of "High" Ratings in the Pre- and Post-Assessment Groups for Each of Five Questions

Group	N	Identify prior learning outcomes. Q1	Identify appropriate assessment methods. Q2	How is that learning related to career goals? Q3	How is that learning related to educational goals? Q4	How is that learning related to personal goals? Q5
Pre-Assessment	104	25%	26%	42%	35%	34%
Post-Assessment	61	34%	40%	55%	52%	51%
Total	165	28%	31%	47%	41%	40%

Question: 14 For those students who have gone through portfolio assessment, does the pattern of responses to the five questions differ from college to college?

Table B4-3 shows the average ratings for the five questions for all students in the post-assessment group. Again, there was a small difference in the overall ratings of the five questions; this may or may not have any practical significance. Also, the average ratings (last column) differed from college to college, but, as previously noted, this seems to reflect different student bodies, not different college "effects." There was not any statistically significant difference in the pattern of ratings for the five questions from college to college (Analysis 4, Appendix 5g). Thus, students do seem to profit from assessment, but there is here no convincing evidence that students are gaining more from some programs than from others or that they are gaining a different type of self-awareness in different programs.

Table B4-3

Pattern of Average Ratings on Specific Questions for Students Who Had Gone Through Portfolio Assessment

Institution	N*	Identify prior learning outcomes. Q1	Identify appropriate assessment methods. Q2	How is that learning related to career goals? Q3	How is that learning related to educational goals? Q4	How is that learning related to personal goals? Q5	Average
B	11	1.9	2.1	2.4	2.4	2.4	2.2
C	4	2.3	2.4	2.4	2.3	2.3	2.3
D	7	2.1	2.2	2.4	2.3	2.4	2.3
F	14	1.8	1.9	2.2	2.3	2.2	2.1
G	7	2.6	2.8	2.9	2.7	2.8	2.8
H	9	1.9	1.9	2.2	2.1	2.3	2.1
I	3	2.2	2.7	2.7	2.7	2.7	2.6
J	6	2.2	2.1	2.2	2.3	2.2	2.2
Total	61	2.0	2.2	2.4	2.4	2.4	2.3

---

*Each average is based upon four 3-point ratings for the number of student papers indicated (3 = high).

Summary

Four studies have been reported in this chapter. Each was introduced as a question. The following paragraphs summarize briefly the answers to those questions insofar as the findings suggest and permit.

- DO FACULTY AGREE ON THE AMOUNT OF CREDIT AWARDED FOR PORTFOLIOS?

In this study, faculty within institutions appeared to agree moderately well in their evaluations of credit for portfolios (both local portfolios and illustrative portfolios). On the other hand, institutions differed widely in the amount of credit they awarded to their local portfolios and the credit they were willing to award the same set of illustrative portfolios. Individual portfolios were often evaluated somewhat differently by different institutions.

- DO FACULTY AGREE ON WHAT CONSTITUTES COLLEGE-LEVEL LEARNING?

The results indicate that there is only a modest degree of agreement among faculty within institutions as to what specific types of learning deserve college credit. Across institutions, some colleges tended to be generally more conservative than others in judging learning to be worthy of college credit, but there was no evidence that colleges differ systematically as to the creditability of specific instances of learning.

- TO WHAT EXTENT DO FACULTY AGREE ON THE TYPE OF DOCUMENTATION REQUIRED AS EVIDENCE OF EXPERIENTIAL LEARNING, AND TO WHAT EXTENT DO VARIOUS TYPES OF DOCUMENTATION PROVE AUTHENTIC?

Colleges appear similar in the extent to which they require documentation though there are some differences as to the general types preferred. Among judges generally there was a clear tendency for the type of documentation required to depend upon the type of learning, but within all institutions there was typically poor agreement among faculty as to whether a particular type of documentation should be required for a specific example of learning. In the second phase of this investigation, ambiguous results and numerous instances of unavailable data made it doubtful that the study deals with the question of authenticity in a way that permits any sound conclusion.

- WHAT IS THE EDUCATIVE VALUE TO THE STUDENT IN DEVELOPING A PORTFOLIO AND HAVING IT ASSESSED?

The results clearly suggest that students who have participated in portfolio assessment exhibit greater self-awareness concerning their learning and their goals than do students who have not participated. There did not appear to be significant differences from college to college in the extent and nature of this positive effect.

Chapter VI

STUDIES OF ASSESSMENT OF WORK COMPETENCES

Amiel T. Sharon and Richard R. Reilly

The relation of work to education has attained national significance in recent years. While educators and citizens alike point to the gap existing between the worlds of work and education, higher education institutions are taking active steps to close the separation. Many individuals are able to work and attend college at the same time. Some drop out of school into work and then drop out of work into school. The formal recognition of knowledge and skill learned by a student on a job is eliminating one of the major barriers to movement from work to school.

Of all types of off-campus learning, the evaluation of work experience is the most prevalent. Thus, one of CAEL's priorities was the development of a model for assessing work experience. This model, called the Work Assessment Model, is described in CAEL Working Paper No. 8. It is designed to help a college identify occupationally specific skills and knowledge, to define the learning outcomes of its occupational and career programs and to translate students' work experience into college credit.

Before any model or assessment procedure can be widely recommended, it is important to research its reliability, validity and other aspects of its utility. Since most of the procedures used for assessing experiential learning require human judgment, there is likely to be more unreliability with these techniques than with the typical written objective examinations. The question of validity is also an important one because the credibility of assessing experiential learning hinges on it. Thus, the purpose of studies of assessment of work competences is to find out whether an accurate assessment of students' job-related knowledge and skill can be made and to evaluate the worth of this learning in terms of college credit.

The participants in the field research on work experience included students and faculty at the following six institutions:

Delaware County Community College
LaGuardia Community College
North Carolina A & T State University
Notre Dame College
Staten Island Community College
Wichita State University

In addition to the students and faculty at the foregoing institutions, the work supervisors of the students also took part in the study.

The general procedure and schedule of the field research is presented in Chapter III. The names of the individuals who coordinated the research at the individual institutions are given in the preface of the report.

Study C1 - DO FACULTY AGREE ON THE AMOUNT OF CREDIT TO BE AWARDED FOR WORK EXPERIENCE?

Description of Study

In order to be fair and equitable, crediting practices must be consistent. One way of determining the consistency of crediting practices is to find out whether different faculty assessors can agree on the number of credits to be awarded to different students after independently examining the same evidence. Thus, the basic purpose of this study was to evaluate the extent to which faculty agree on the number of credits awarded for prior work experience.

A total of 90 students attending four colleges were the subjects of this study. There were 46 students seeking credit in the field of Law Enforcement, 27 in Secretarial Science, 9 in Data Processing, and 8 in Management. Table C1-1 shows how the students in the three occupational fields were distributed among the four institutions.

Each student was assessed independently by two faculty judges who are experts in the subject matter in which the student was seeking credit. The specific method of assessing work experience varied among institutions but all used one or more parts of the Work Assessment Model. Most of the institutions used the task inventory checklists (presented in Appendix 4x and in the Appendix of Working Paper No. 8) as the starting point in their assessment process.

The assessors in this study were the faculty members selected from departments teaching the subjects in which the assessment was conducted; i.e., Business Management, Law Enforcement, Data Processing and Secretarial Science. At two of the four participating colleges only about half of the

Table C1-1

Number of Students Awarded Credit for Prior Work Experience by Two Faculty Assessors at Different Institutions In Three Occupational Subjects

Institution	Management	Law Enforcement	Data Processing	Secretarial Science	Total
A	0	17	0	0	17
B	8	0	3	7	18
C	0	9	6	0	15
G	0	20	0	20	40
Total	8	46	9	27	90

faculty had prior experience in assessment of experiential learning. All were experienced in assessment at the third college and none had prior experience at the fourth.

Preparation of the faculty to conduct assessment was minimal. Each assessor was given Working Paper No. 8 and the instructions and forms for recording the amount of credit awarded to each student (see Appendix 4u for the instructions and forms). No attempts were made to hold workshops or other kinds of training sessions prior to assessment.

Based on his or her assessment, each faculty member was requested to assign credit to a student in each of 7 pre-defined content areas within the occupational subject matter. For example, two content categories within Law Enforcement are "Police Operations" and "Criminal Investigations." The total number of credits assigned to a student was the sum of the credits in each of the content categories.

At two of the participating colleges the students were assessed as part of the institutions' regular assessment program. At the other two colleges the students were randomly selected from a pool of students who were known to have experience in a particular occupation. At one of these two colleges it became necessary to compensate students in order to encourage participation in the study.

Data Analysis

The dependent variable or the basic units of analysis were the number of credits assigned to each student within each content category and the total number of credits assigned to students by pairs of raters was analyzed within and between institutions, occupational fields, and content categories within fields. This variation provided estimates of interrater agreement or reliability. A technical discussion of the data analysis and reliability is presented in Appendix 5h.

Results

Question: How much credit is awarded for work experience?
1

Table C1-2 indicates the average amount of credit awarded students for work experience by institution and occupational field. The overall average is 8.4 credits and is equivalent to almost three 3-credit courses. Although there is a highly significant difference in the amount of credit awarded in different occupational curricula, the interpretation of this result is unclear. Since students in different fields are in different colleges, the difference in credit could be an institutional effect. Although Table C1-2 does suggest the possibility that students in Law Enforcement acquire more college-equivalent knowledge than those in the other fields considered, this interpretation must remain mere speculation for the time being.

Question: What is the consistency of judgment in assigning credit to a specific content category of occupational subject matter?
2

Since there are differences in the types of competences acquired in different occupations, one might hypothesize that the reliability of awarding credit depends upon the occupational subject matter for which credit is awarded. Table C1-3 illustrates the data from one institution regarding this question. The table indicates the number of credits assigned by two faculty assessors to eight students in the field of Law Enforcement. Two types of error or inconsistency are illustrated in the table. One kind of error results from the tendency of assessor A to

Table C1-2

Average Number of Credits Awarded Students for Prior Work Experience by Institution and Occupational Field

Institution	Occupational Field				
	Management	Law Enforcement	Data Processing	Secretarial Science	All Fields
A		14			14.0
B	2		3	7	4.1
C		13	4		9.4
G		12		3	7.5
All Institutions	2	13	3.6	4	8.4

Table C1-3

Illustrative Data Showing Number of Credits Awarded for Prior Work Experience in Law Enforcement by Two Faculty Assessors (A and B) to Eight Students at One Institution

Student Code	Content Category														Total Number of Credits	
	Police Operations		Criminal Investigation		Criminal & Const. Law		Criminal Evidence		Criminal Procedures		Administration		Other Activities			
	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
1	3	1	3	1	3	1	3	1	3	1	0	1	3	3	18	9
2	3	1	3	1	3	1	3	1	3	1	3	3	9	3	27	10
3	3	0	3	0	3	0	0	0	3	0	0	0	0	1	9	1
4	3	2	3	1	0	0	3	1.5	3	1.5	3	3	9	3	27	12
5	3	1	3	1	3	0	3	1	3	1	3	3	9	3	27	10
6	3	1	0	0	0	0	0	0	0	0	0	1	3	0	6	2
7	3	1.5	3	1.5	3	0	3	1.5	3	1.5	3	3	9	2	27	11
8	3	1	3	1	3	0	3	1	3	1	3	3	9	3	27	10

assign more credit than assessor B to all students in almost all content categories. The second kind of error is that the relative difference between the two raters is not constant across students. For example, under the content category of "Police Operations" the two raters differed by 3 credits with respect to Student 3 while they differed only by 1 credit with respect to Student 4.

In order to answer the question whether the faculty assessors agree on the number of credits to be assigned within each category of an occupational subject matter, analysis was conducted within each institution and occupational field where the number of students was greater than 10. Specifically, this analysis was conducted in Institution A (Law Enforcement) and Institution G (Law Enforcement and Secretarial Science). Each of these analyses suggests the assessors do not agree in their discriminations when granting credit among the various content categories. Because of the small number of cases in each of these analyses, the results are not conslusive regarding this question.

Question: 3 Are there differences among institutions and occupational fields in the consistency of faculty agreement when credit is awarded to students?

Table C1-4 indicates the reliabilities (consistency of faculty agreement) for each institution-occupation combination. Two kinds of reliability coefficients were computed: interrater agreement and interrater consistency. The former coefficient takes into account all of the discrepancies between credit assignments by pairs of assessors while the latter focuses primarily on the consistency of rank-ordering students on the basis of the number of credits. (A technical discussion of the difference between interrater agreement and consistency is presented in Appendix 5h.) It can be seen that there is a wide range of reliabilities. In certain departments,

Table C1-4

Interrater Agreement and Consistency Coefficients Indicating the Reliability of Awarding Credit for Prior Work Experience by Institution and Occupational Field

Institution	Department	N	Interrater Agreement	Interrater Consistency
A	Law Enforcement	17	.00	.80
B	Management	8	.86	.85
B	Data Processing	3	.96	.96
B	Secretarial Science	7	.00	.56
C	Law Enforcement	9	.00	.00
C	Data Processing	6	.91	.93
G	Law Enforcement	20	.00	.72
G	Secretarial Science	20	.69	.70
Overall Median			.35	.76

particularly in Law Enforcement, the interrater agreement is low but the consistency is high. This result suggests that training the faculty assessors on how much credit to award is likely to increase reliability. Although the reliabilities in the fields of Management and Data Processing are fairly high, the small number of cases in these fields tends to limit the generalizability of the results. The overall median coefficients which are likely to be more stable do suggest, however, that there is only modest interrater agreement in assigning credit while the interrater consistency is quite acceptable.

Question: 4 What are the absolute differences in the amount of credit awarded by two faculty assessors of each student?

Another way to evaluate the consistency of judgments of two evaluators is by determining the absolute difference between their judgments. For each pair of judgments it is possible to compute the difference, if any, and to set up a frequency distribution of such differences. Table C1-5 presents the frequency and percentage of pairs of judgments differing by a specific number of credits. The median difference for all pairs of faculty assessors in all fields was 4 credits. Examination of Table C1-5, however, reveals that most of the differences occur in Law Enforcement, and most of the differences appear to result from large differences in the average amount of credit awarded by the different raters. There are mostly small or no credit differences in the other three fields. The median difference in Law Enforcement was 18 while it was 0 in Management, 1 in Data Processing, and 1 in Secretarial Science.

Table C1-5

Number and Percentage of Faculty Assessors Differing by a Specific Number of Credits Awarded in Four Areas of Prior Work Experience

Difference in Credits Awarded	Management		Law Enforcement		Data Processing		Secretarial Science		All Fields	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
0	5	62	0	0	2	22	8	30	15	17
1	3	38	3	7	3	33	7	26	16	18
2			0	0	4	44	4	15	8	9
3			1	2			2	7	3	3
4			1	2			4	15	5	6
5			1	2			0	0	1	1
6			2	4			0	0	2	2
7			0	0			1	4	1	1
8			1	2			0	0	1	1
9			3	7			1	4	4	4
10			0	0					0	0
More than 10			34	74					34	38
Total	8	100	46	100	9	100	27	100	90	100

Study C2 - TO WHAT EXTENT CAN FACULTY AGREE ON ASSIGNING CREDIT FOR WORK EXPERIENCE TO COURSES WITHIN A PARTICULAR OCCUPATIONAL PROGRAM?

Description of Study

The Work Assessment Model suggests that the evaluation of prior work experience be conducted in the context of the courses in the relevant occupational program. The model assumes that it is possible to identify specific learning outcomes of work experience and to translate them into courses or other units of the curriculum. The major purpose of this study was to find out the degree to which faculty within an institution can agree on the relevance of different competences to specific content categories within an occupational field.

Instead of assessing actual students in this study, illustrative portfolios of actual students were used. The use of illustrative portfolios made it possible to obtain comparable data from different institutions. Five different portfolios in each of the areas of Management, Law Enforcement, and Data Processing were evaluated by two or three faculty raters at six institutions (see Appendix 4v for an illustrative portfolio). Table C2-1 indicates the number of faculty assessors within each institution and occupational field.

The faculty assessors had varied backgrounds. Most had some practical experience in the field that they were evaluating. Some had only teaching experience. The only preparation given to most of the assessors was an explanation by the campus research coordinator on how the illustrative cases were to be assessed.

Table C2-1

Number of Faculty Assessors Awarding Credit to Each of Five Illustrative Cases by Institution and Occupational Field

Institution	Management	Law Enforcement	Data Processing	Total
A	3	3	3	9
B	2		3	5
C	3	2	3	8
D	3	2	3	8
E	2	2	3	7
G		2		2
Total	13	11	15	39

One of the apparent problems that emerged with the use of the illustrative portfolios is that several assessors felt that there was insufficient documentation for some of the portfolios. Some of the faculty may have assigned little or no credit to college-equivalent learning primarily because adequate verification was not present. The question of what constitutes adequate verification, however, is part of the assessment problem. The extent to which there is agreement or disagreement on this question is likely to influence the reliability of assessment.

Data Analysis

As in Study C1, the dependent variable or the basic units of analysis was the number of credits assigned to each case in each content category within an occupational subject matter.

Table C2-2 presents sample data used in this study showing the total amount of credit awarded in one subject matter at one institution. The table presents the total number of credits only, but does not show the content category credits which sum to the total number.

Table C2-2 indicates that the average case at this college was assigned 9 credits. Assessors X and Y are in very close agreement regarding the total number of credits to be granted to the illustrative cases. On the other hand, Rater Z assigned much less credit to the illustrative cases than the other two raters. All three assessors rank-ordered the cases in about the same way. These data suggest that the interrater agreement is relatively low and the interrater consistency is relatively high.

A detailed discussion of the data analysis used in Study C2 and a full explanation of the difference between interrater agreement and consistency is given in Appendix 5i.

Table C2-2

Illustrative Data Showing Total Credit Awarded for Prior Work Experience in Data Processing in One Institution

(Five Illustrative Cases Rated by Each Faculty Assessor)

Illustrative Case	Faculty Assessor X	Faculty Assessor Y	Faculty Assessor Z	Average
1	21	15	7	14.3
2	9	9	3	7.0
3	3	9	1	4.3
4	18	15	1	11.3
5	9	12	3	8.0
Average	12	12	3	9.0

Results

Question: 5 How much credit is awarded to illustrative cases for prior work experience?

Table C2-3 indicates the average amount of credit awarded to illustrative cases by institution and occupational field (the number of credits in the table were rounded to the nearest whole number). The overall average number of credits awarded was 12. This number of credits is equivalent to four 3-credit courses and is somewhat greater than the average number of credits awarded to actual students in Study C1 (which was 8.4 credits). Any comparison between the two studies, however, could hardly be conclusive because different groups of students were assessed in each.

It will be recalled that in Study C1 significant differences were found between the occupational fields, but no conclusions could be drawn about the finding because different institutions were assessing students in different fields and different students in the same fields. In this study, however, all institutions (except B and G) assessed the same students in the same fields. It is, therefore, interesting to note that although both institutional and field differences in the number of credits awarded remain, they are relatively smaller than those in Study C1.

Question: 6 To what extent do faculty agree on the amount of credit awarded to illustrative cases?

Within each of the 15 departments that participated in this study an analysis of the interrater agreement and consistency was conducted. Table C2-4 indicates the reliabilities found in each of the departments. It can be noted that the reliability coefficients range from a low of .00 to a high of .87. Of particular interest is the finding that in

Table C2-3

Average Amount of Credit Awarded to Illustrative Cases for Prior Work Experience in Three Occupational Fields

	Occupational Field			
Institution	Management	Law Enforcement	Data Processing	All Fields
A	28	15	9	17
B	6		7	6
C	12	18	12	14
D	17	10	10	13
E	9	11	8	10
G		12		12
All Institutions	14	13	9	12

6 of the 15 departments there is a substantial difference between the two kinds of reliability. This finding suggests that the general tendency by assessors to award a small or a large number of credits is a major reason for poor interrater agreement.

Since the reliabilities were computed only on a small number of cases within each department (2 assessors evaluated 5 portfolios) the individual coefficients are not very meaningful. The pattern of the reliabilities is more generalizable as are the overall median coefficients presented in the table. These medians indicate that although the agreement between assessors is low, the consistency is high. By training the assessors how to evaluate different kinds of learning in terms of amount of credit normally awarded, it is very likely that agreement would be improved.

As was done with actual students in Study C1, the reliability of judgment was also evaluated by determining the absolute difference in the number of credits assigned by two assessors to the same student. Again, for each pair of judgments the difference, if any, was computed and a frequency distribution of such differences was set up. Table C2-5 presents the frequency and percentage of pairs of judgments which differ by a specific number of credits. The difference in credits are those which are awarded to one illustrative case. Since in each field there were 5 illustrative cases and 5 institutions, a total of 25 pairs of judgments were made in each occupational field.

The median difference between all pairs of judgments was 6 credits. There were, however, significant differences among the three occupational fields. The median difference in Data Processing was 4 credits, while it was 6 credits in Management and more than 10 credits in Law Enforcement.

Table C2-4

Interrater Agreement and Consistency Found in 15 Different Departments Showing Whether Faculty Tend to Agree on Amount of Credit to be Awarded to Illustrative Cases

Institution	Department	Interrater Agreement	Interrater Consistency
A	Management	.00	.47
A	Law Enforcement	.69	.87
A	Data Processing	.12	.67
B	Management	.77	.71
B	Data Processing	.63	.64
C	Management	.34	.73
C	Law Enforcement	.00	.03
C	Data Processing	.83	.90
D	Management	.00	.48
D	Data Processing	.00	.48
D	Law Enforcement	.00	.19
F	Management	.00	.00
F	Law Enforcement	.62	.67
F	Data Processing	.76	.86
G	Law Enforcement	.00	.65
	Median	.12	.65

Table C2-5

Number and Percentage of Faculty Assessors' Judgments Differing by Specific Numbers of Credits Awarded to Illustrative Cases for Prior Work Experience by Occupational Field

Difference in Credits Awarded	Occupational Field							
	Management		Law Enforcement		Data Processing		All Fields	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
0	2	8	3	12	5	20	10	13
1	3	12	0	0	2	8	5	7
2	2	8	0	0	0	0	2	3
3	4	16	2	8	4	16	10	13
4	1	4	0	0	2	8	3	4
5	0	0	0	0	3	12	3	4
6	4	16	3	12	3	12	10	13
7	3	12	0	0	0	0	3	4
8	1	4	0	0	0	0	1	1
9	1	4	2	8	3	12	6	8
10	0	0	2	8	1	4	3	4
More than 10	4	16	13	52	2	8	19	25
Total	25	100	25	100	25	100	75	100

Question: 7 To what extent do faculty assessors agree on assigning credit to a specific content category of occupational subject matter?

This question is basic to Study C2 for the answer indicates whether faculty assessors can make discriminations among different content categories in a subject matter and assign credit accordingly. Agreement among the assessors on the breakdown of credit into the seven content categories in each field would be indicated by a consistent tendency for some students to be assigned more credit in certain content categories than in others. This would also indicate that the faculty assessors can make discriminations among the content categories.

Table C2-6 summarizes the results of the analyses which show whether faculty tend to agree on the breakdown of credit. Since only 3 of the 15 institution-field combinations show significant faculty agreement, the results are hardly conclusive. The overall pooled results do suggest, however, that credit can be broken down reliably in Data Processing but not in the other two occupational fields.

Table C2-6

Results of Analyses Showing Whether Faculty Assessors Tend to Agree on the Breakdown of Credit Awarded in Different Content Categories Within an Occupational Subject Matter

Institution	Management	Law Enforcement	Data Processing
A	No	No	No
B	Yes		No
C	No	No	Yes
D	No	No	No
E	No	No	Yes
G		No	
Overall	No	No	Yes

"Yes" means there was a statistically significant degree of agreement between the faculty assessors on the breakdown of credit.

Study C3 - WHAT IS THE CONCURRENT VALIDITY OF THE DOCUMENTATION OF WORK EXPERIENCE?

Description of Study

A basic element of the Work Assessment Model is the inventory of work competences. In order to be valid, there should be an independent corroboration of the student's statements regarding which activities have been conducted on the job. The purpose of this study was to determine the extent to which students and their work supervisors agree on which specific activities were learned by the students on the job.

The task inventory checklists in Law Enforcement, Data Processing, and Secretarial Science (taken directly from Working Paper No. 8) were completed by 72 students and their work supervisors. Only at one of the five institutions that participated in this study were the data collected as part of the regular institutional assessment process. At the remaining colleges the data were collected only for study purposes. Table C3-1 indicates the number of student-supervisor pairs within each institution and occupational field.

Data Analysis

The data were analyzed by computing the frequency and percentage of checklist tasks checked by the student and his or her work supervisor. Since the checklist tasks were grouped into content categories, it was possible to compute the frequency of student and supervisor checks within a content category. The details of the analyses are described in Appendix 5j.

Table C3-1

Number of Student-Supervisor Pairs Who Completed Task Inventory Checklists by Institution and Occupational Field

	Occupational Field			
Institution	Law Enforcement	Data Processing	Secretarial Science	Total
A	8			8
B		7	7	14
C	9	6		15
D	1	1	2	4
G	17		14	31
Total	35	14	23	72

Results

Question: 8 To what extent do students and their work supervisors agree or disagree on the number of tasks learned by the student on the job?

One way to evaluate the degree of agreement between students and their work supervisors is by comparing the number of tasks checked within each content category of an occupation, regardless of whether the tasks checked are the same. For example, if a Data Processing student checked 15 out of 158 tasks in the content category of "Computer Programming" and his supervisor checked 15 different but possibly similar tasks in the same content category, their agreement would be considered perfect for this particular analysis. The reasoning for this type of scoring is that the content categories are fairly homogeneous and just like a test score, it is possible to obtain the same score by different combinations of items.

Since each checklist contains six sections corresponding to content categories, it was possible to compute six scores for each student and each work supervisor.

The task inventory checklists used in this study contain lists of a large number of tasks. There are a total of 139 tasks in Law Enforcement, 330 in Data Processing and 157 in Secretarial Science. Since these lists provide a comprehensive coverage of the activities performed in a field, it was not expected that any individual will have accomplished all or even most of the tasks listed. Table C3-2 presents the average number of checklist tasks checked by students and their work supervisors by institution and occupational field.

Table C3-2 indicates that, in general, students say that they have performed slightly more tasks than their supervisors say. There are,

Table C3-2

Average Number of Checklist Tasks Checked by Students and Their Work Supervisors by Institution and Occupational Field

Institution	Law Enforcement		Data Processing		Secretarial Science		Average for Each Institution	
	Students	Supervisors	Students	Supervisors	Students	Supervisors	Students	Supervisors
A	153	158					153	158
B			27	23	46	64	34	46
C	114	125	65	34			97	95
D	150	157	36	36	84	94	89	95
G	121	81			37	37	82	61
Average for Each Field	127	112	44	29	42	50	84	76

however, some significant differences among different departments with respect to whether students or supervisors check more tasks. Specifically, of the 10 departments (institution-field combinations) involved in this study, in 5 departments the supervisors checked more tasks than did the students.

Figure 1 presents the same kind of data as Table C3-2 but broken down by content category within a field across all five institutions. Figure 1 is based on the data from Table C3-3. The graphs in the figure indicate that there is a close relationship between the average number of tasks checked in each content category by the students and their supervisors.

Question: To what extent are students' claims to have conducted
9 specific tasks verified by their supervisors?

Another way to evaluate the degree of agreement between students and their supervisors is by the number of tasks checked by the student which are also checked by his or her supervisor. Table C3-3 indicates the total number of tasks checked by students and the percentage of verified tasks by subject matter and content category. The percentage of verified tasks was computed by using, as the base, the total number of tasks checked by the student within each content category. Thus, if a student checked 10 tasks within a particular content category and his supervisor also checked 10 tasks, only 5 of which were the same tasks, then 50 percent of the tasks are verified.

It can be seen in Table C3-3 that in almost all content categories in Law Enforcement and Secretarial Science a substantial proportion of the tasks have been verified. On the other hand, in Data Processing only in one of the six content categories are more than half of the tasks verified.

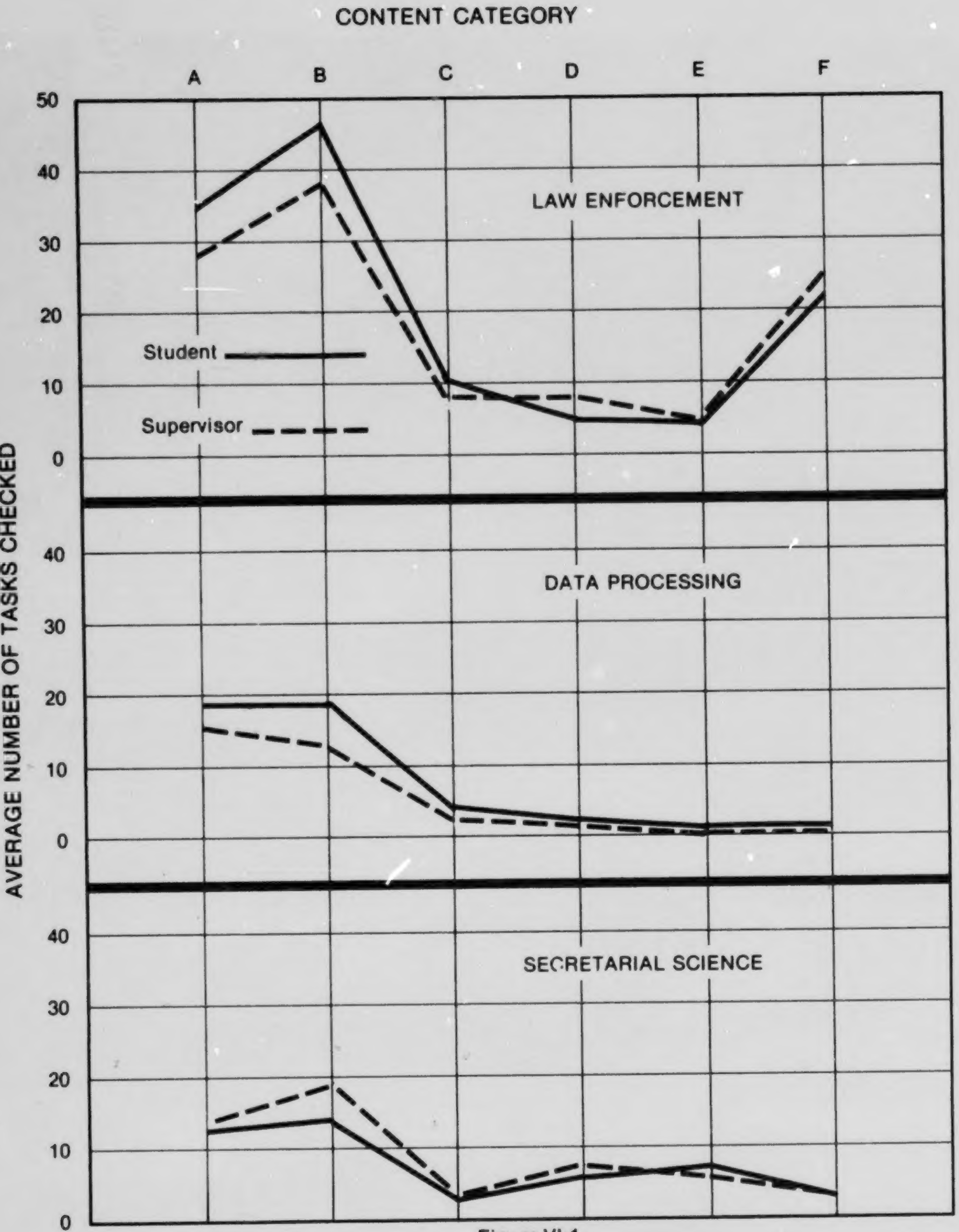


Figure VI-1

Average Number of Checklist Tasks Checked by Students and Their Supervisors at Five Institutions by Content Category in Three Occupational Fields

Table C3-3

Total Number of Tasks Checked by Students and Percentage of Those Tasks Which Are Verified by the Supervisors in Different Occupational Fields and Content Categories

Content Category	Law Enforcement		Data Processing		Secretarial Science	
	Number	Percent[a]	Number	Percent[a]	Number	Percent[a]
A	1223	67	263	68	269	85
B	1646	72	265	38	305	72
C	345	83	37	5	67	55
D	257	79	24	0	144	70
E	251	81	17	6	151	57
F	741	67	12	0	36	39
Total for Each Field	4463	71	618	46	972	71

[a] Percent = $\frac{\text{Number of Verified Checks}}{\text{Total Number of Student Checks}}$

The total percentages of verified tasks across all content categories and fields do suggest that work supervisors corroborate students' claims regarding what was done on the job. Even in Data Processing approximately one-half of the tasks are corroborated.

Summary

It might be best to summarize the results of the three studies in terms of the steps or phases suggested by the Work Assessment Model. One of the initial steps in the model is to identify and verify the student's work competences. Thus, a basic question that the field research was designed to answer is: what is the concurrent validity of the documentation of work experience?

The results of Study C3 strongly suggest that students' work supervisors corroborate students' claims regarding activities on the job. There is a good reason to believe that task inventory checklists by which the student describes his or her work experiences yield accurate and authentic information.

After verifying the student's work competences it is necessary to relate them to the curriculum. Since the curriculum is structured by content areas, work competences should be translated into these areas when awarding credit. Thus, a second basic question which the field research was designed to answer is: to what extent can faculty agree on assigning credit for work experience to courses within a particular occupational program?

The results of Studies C1 and C2 are inconclusive regarding the extent to which faculty assessors can agree on the assignment of credit to different content categories in a particular subject matter. The number of cases used in these analyses were too few for confident generalization and the results were conflicting, but the findings provide only limited evidence of consistent articulation of occupational competence to subject matter categories.

The final step in the Work Assessment Model is to award credit to the student. Regardless of what was done prior to this step, it is important that the number of credits awarded represent a reliable and fair standard of learning or competence. Can faculty agree on the amount of credit that should be awarded for work experience to different students?

The results of Studies C1 and C2 indicate, with some qualification, that faculty assessors tend to be consistent with each other in their credit assignments. Although some assessors are more liberal or conservative in awarding credit, there is substantial agreement among them on which students should be granted more or less credit. The training of such assessors can remove much of the remaining disagreement between them.

Chapter VII

SELECTED INSTITUTIONAL STUDIES OF EXPERT JUDGMENT

Alverno College - Georgine Loacker
Antioch College - Ruth Churchill
Thomas A. Edison College - Paul I. Jacobs

INTRODUCTION

Warren W. Willingham

In addition to the need to examine assessment assumptions and practices as reported in the previous chapters, CAEL also recognized the need to encourage the development of routine steps for monitoring the quality of assessment procedures. It was felt that a useful first step would be to facilitate local studies of expert judgment that were specifically focused upon the measurement adequacy of operating assessment procedures.

Accordingly, one option for institutional participation in CAEL Field Research was to carry out an institutional study. Richard R. Reilly, an ETS staff member, developed for this purpose a prototype study outline to serve as a general guide (see Appendix 4y). Using that prototype as a model but directing their attention to the character of their own programs, ten institutions did carry out such a study and submitted reports based on their work.

The Executive Committee reviewed these reports and considered which might be especially useful to other institutions. The following three reports printed here were chosen not only because they were judged to be worthwhile examples but because they were each somewhat different in subject matter and approach. While CAEL recommends them as interesting and useful studies, they are the work of the authors and their institutions and, like all CAEL institutional reports, may reflect practices different from those suggested elsewhere by CAEL.

STUDY D1:

A RELIABILITY STUDY OF EXPERT JUDGMENT IN THE ASSESSMENT OF SPEAKING PERFORMANCE OF ENTERING COLLEGE STUDENTS

Georgine Loacker, Alverno College

Program Within Which the Assessment Is Carried Out

The assessment procedure chosen for this reliability study is one that has been in use at Alverno College for three years. It involves the assessment of the speaking performance of entering students and is administered before the student begins her college work. This appraisal, along with that of writing, listening, reading, and graphing, constitutes the first-level assessment of the communications competence, which is one of eight competences that each Alverno student must demonstrate at a specified level in order to fulfill degree requirements.

At this first level of competence, the student must demonstrate the ability to identify her own strengths and weaknesses in areas of communications. Thus she begins a process of self-assessment which she develops throughout her academic career, as part of the total process of becoming a lifelong independent learner.

In the first-level communications assessment, of which the speaking assessment is a part, the student's performance is evaluated by a team of two expert judges. An important aspect of the team assessment is the consensus process. After the assessors each record student behavior and make an independent judgment on the basis of their observations, they meet together with the explicit goal of reaching consensus. They compare their judgments and create a composite of behavioral evidence. They examine the composite until they agree on the judgment for each criterion. An important assumption underlying the process is that multiple perceptions can ultimately provide a more complete and richer composite of the student's behavior as a basis for understanding and evaluating her performance.

Later, during the first week of the term when the student has some distance from her performance, she evaluates it herself. Then in a feedback session she restudies her own behavior with the evidence and analysis provided by her assessors. She works until she is successful. Following essentially the same process as the assessors, she learns, from the person who provides feedback, how to observe her own behavior precisely and completely enough to judge whether or not it meets the criteria. She is judged successful for this level of competence when she makes 5 out of 6 judgments that the assessors have agreed are supported by the behavioral evidence available. She may negotiate the assessors' judgment only if she can change the composite by additional significant behavioral observation that would result in new judgment.

Specific Assessment Procedure

The students whose speaking assessments were used for this study were 34 students who were among those entering Alverno in January, 1976. They were assessed during the January interim, the first two weeks of that month. Their ages range from 20 to 55.

Participants were given Part I of one of the forms of the level-one speaking instrument now in use (see Appendix 6a). This consists of a set of directives telling them to prepare and present a speech on a specified topic to a specified audience. In addition to the directions written on the student's copy of the instrument, the person who administered the test reinforced the important aspects (see Appendix 6b).

The topic chosen for the speech was one that would enable students to express ideas, organize them, and marshal evidence from their own experience so that they could demonstrate their speaking ability apart from their academic knowledge. In order to eliminate one source of anxiety, no strict time limits were set for the preparation period. Each student's speech was videotaped so that she could carefully view her performance later in order to evaluate it. Consequently, it was not necessary for the assessors to be present when the student delivered her speech. The two expert judges for the present study assessed the 34 student speeches from cassette videotapes in two afternoon sessions.

The assessors were two college graduates with BA's in Speech. Prior to this assessment session they had worked together as a team on one other set of about 150 student speeches. They had also judged three or four sets with other trained assessors during the past three years. Their original training consisted in careful study of an assessment manual (see Appendix 6c), a half-day presentation and laboratory session, and the equivalent of two days of laboratory practice sessions. In these practice sessions they received feedback on the evidence they recorded as the basis for their judgments and on the correlation of their judgments with those of other assessors.

The criteria for judging the student speeches (see Appendix 6d) were established by two college instructors, one with an advanced degree in Speech and one in Communications. In a process of trial and refinement, these criteria have been reduced from an original nine to the present six. They were expressed in behavioral definitions--with further specification of behavioral examples in the assessment manual (Appendix 6c).

In the assessment process, the two experts independently made their decisions on each criterion (Appendix 6d) and recorded relevant evidence or explanation. They made no judgment on the speech as a whole since the point of the assessment is identification of the student's strengths and weaknesses.

After this first phase the assessors met and reached consensus on each criterion (see Appendix 6e). Finally, each student judged her own speech on the basis of the same criteria (Appendix 6f). In instructional feedback sessions she studied her own speech and the criteria applied until she learned to apply the criteria and perceive evidence. She was credentialized for this level of communications when five out of six of the strengths and/or weaknesses she identified correlated with those identified by consensus of the assessors.

Data Collection

The following data represent the independent assessments of two different trained judges, as well as the self-assessment of the students, on 34 student speeches. Each judge separately viewed videotapes of the speeches and made a judgment whether or not each speaker met each of six criteria, at the same time recording evidence and explanation for their judgments. Each judgment was recorded as a yes (strength of the speaker) or no (weakness of the speaker).

Results of Data Analysis

Since there are no actual "scores," results are reported in terms of agreement on items. Table D1-1 lists the number of strengths and/or weaknesses on which the two trained judges were in agreement for each speaker.

Table D1-1

Number of Strengths and/or Weaknesses Agreed Upon
By Both Trained Judges for Each Speaker
(Total items = 6)

Student #	No. of Items	Student #	No. of Items	Student #	No. of Items
1	5	13	5	24	4
2	5	14	6	25	4
3	4	15	5	26	6
4	6	16	5	27	5
5	5	17	6	28	4
6	6	18	4	29	5
7	4	19	5	30	6
8	5	20	5	31	6
9	5	21	6	32	6
10	5	22	6	33	6
11	6	23	4	34	6
12	6				

Grouped Distribution of the Extent of Agreement Data

Number of perfect agreements on all criteria (6 out of 6)	14
Number of agreements on 5 out of 6 criteria	13
Number of agreements on 4 out of 6 criteria	7

Table D1-2 shows the correlation between the two trained assessors for each of the six criteria, the reliability of the composite judgment of the two raters for those criteria, and the correlation of student self-assessments with the composite judgments. The first two sets of coefficients are taken as strong support for high agreement between the ratings of trained assessors. The final column of coefficients constitutes partial evidence of the yet undeveloped skill of self-assessment in entering students.

Conclusions

The level of agreement between the two trained judges suggests a high degree of reliability for the individual expert judgments involved. Within the present developmental stage of assessment in the Alverno curriculum, this study verifies the existence and significance of both the careful specification of explicit criteria and the training of assessors.

The low level agreement between the students' judgment of their own performance and the composite judgment of the two trained assessors supports an important assumption of the learning program at Alverno: Entering college students have very limited ability to assess their own performance accurately; designed learning experiences are necessary for them to develop their ability to observe their own behavior carefully and make judgments out of that observation.

The study also raises a series of questions for further studies to probe. For which individual assessments, for example, would one assessor be sufficient if he/she is carefully trained and the criteria clearly specified? Because the Alverno assessment process is built on a principle of multiple instruments and multiple assessors, careful collection of

Table D1-2

Correlations and Reliability Coefficients for the Criteria

	Correlation Between Two Ratings	Reliability of the Composite of Two Ratings	Correlation Between Self-Assessment and Composite Rating
Criterion 1	.81	.90	.21
Criterion 2	***	***	.25
Criterion 3	.87	.93	.31
Criterion 4	.53	.69	.35
Criterion 5	.77	.87	.23
Criterion 6	.33	.50	.25

*** Connotes that the phi coefficient is indeterminant

✓

data is needed to clarify the specific meaning and application of the principle in different assessment situations which are part of the whole.

What can we learn about assessors, procedures, and instruments from comparing individual judgments with final team consensus? From examining the frequency with which items are judged as strengths? From determining which items tend to bring forth agreement most frequently? From observing, within assessor judgments, patterns of frequency in judging strengths and/or weaknesses? All of these questions are relevant to a general plan of ongoing evaluation as a means of development and quality control for the assessment process at Alverno. The present study provides a starting point.

STUDY D2:

DEVELOPMENT OF A COMPETENCY-BASED SCALE FOR WRITING

Ruth Churchill,* Antioch College

Introduction

For many years the Yellow Springs campus of Antioch College has required that all students demonstrate competency in writing as a requirement for entering the senior year. The assessment of this competency has been made in a variety of ways. Since 1956 there has been no required course in composition. Competenc in writing is demonstrated in two ways: first, by a satisfactory grade on an essay at entrance; second, by satisfactory writing in courses up to the senior year. Students whose writing is identified as not meeting standards at entrance are advised to seek help, which is available in courses emphasizing writing. Students whose writing in courses is unsatisfactory have to work with a tutor until the tutor is satisfied with their written work.

From 1956 to 1970 the procedure was for all entering students to write a short essay during the orientation period. Essays were rated by teams of two independent readers, each reader judging a group of 30 to 40 papers. The readers were selected from a panel of educated adults--not members of the college faculty, but individuals who had experience related to writing or to teaching writing. Ratings were made in terms of six criteria and an overall judgment. Each reader's ratings were converted into standard scores to avoid the effects of systematic differences in severeness of judgment by different readers. In case of serious disagreement by readers, the essay was read by a review reader. Since this system was normative, it had the disadvantage that the standards for good, average, and poor performance were based on the year's entrants or (using techniques for tying one year to another) several year's entrants. Some performances had to be good; others had to be poor.

---

* With the assistance of Peggy Brown in the collection and analysis of the data.

From 1971 to 1974 it was assumed that most students admitted to Antioch were already writing at an acceptable college level. It was expected that good students with writing disabilities and students with borderline writing would be alerted to their problem by course instructors. The task was to identify at entrance students needing work; all students scoring less than 475 on the verbal section of the College Entrance Examination Board Scholastic Aptitude Test were asked to write a short essay. An attempt was made to approximate previous standards by using graders and topics previously used and employing past distributions of standard scores for these graders and topics. This approach proved even less satisfactory than the preceding one. Students felt singled out as deficient, many students who needed help were not identified, and it was impossible to maintain the same standards as in previous years because graders were more influenced by the current group of students than by their past standards.

Thus, in 1975 it was decided to use a competency-based rather than a normative approach to assessing performance in writing; after all, competency in writing was the graduation requirement. Faculty committees working on the program for entering students felt that the new competency approach, requiring a sample of writing of all entering students followed by attention to writing in courses, was an important part of the program planned.

The overall purpose of the research reported here was the development of a competency-based scale for assessing performance in writing. The first step involved the development and the pretesting of the scale. The second step involved assessing the extent to which raters could make similar judgments with the scale. The third, and final, step was to follow up a subsample of students initially rated with the scale, using their performance in writing in their courses as a criterion.

Definition of Competency in Writing

The first problem was to define what is meant by competency in writing. It was decided to define not college-level writing but the level at which an educated adult needs to write. The goal of college is not to prepare students who can write satisfactory papers in courses but adults who can meet daily demands for writing, such as a business letter or a report to a committee. An educated adult should be able to write in a clear and organized fashion with only minor errors in grammar and spelling. Of course, some individuals can write with style, even when they enter college, but publishable writing is not demanded of all educated adults. On the other hand, major grammatical errors, misuse of words, inability to express ideas, and lack of organization indicate a need for developing greater skill in writing.

Development of the Scale

With this general definition in mind, it was possible to devise a competency-based scale and to select readers. Two sources were available for the scale: first, the previous grading sheet (see Appendix 6g) and the statement of standards for writing on which it was based (Appendix 6h), and second, the behaviorally-anchored scale for writing competence published in CAEL Working Paper No. 10 (page 16).* The scale embodies the standards that the faculty agreed on in two columns, one emphasizing clarity of statement, appropriate use of vocabulary, and correct grammar (including spelling and punctuation), and the other stressing organization, expression of ideas, and style. Each level of writing is defined in terms of both columns. While five levels of writing are indicated, there are nine points on the scale. The additional four points represent intermediate positions, the need for which readers indicated after a trial run. Two points on the scale represent one level in writing. (See Appendix 6i).

---

*CAEL Working Paper No. 10 will be published in revised form in early 1977 as Expert Assessment of Experiential Learning--A CAEL Handbook, by Richard R. Reilly, et al.

The choice of readers followed both from the definition of competency in writing and from past experience. If the level of competency in writing demanded is that expected of an educated adult, then the best experts are educated adults. Past experience in using readers with backgrounds related to writing has already indicated that such readers made judgments and comments useful to faculty and students. Therefore, the readers continued to be educated adults with experience in writing or in teaching writing. Since communication involves both the writer and the reader, it has always seemed wise to allow for individual idiosyncracies on both sides by having all papers read by two individuals and reviewed by a third if the first two disagreed seriously (more than one level or two points on the scale).

(A note on the cost of this procedure may be appropriate. It is not expensive. One reader can handle 30 to 40 papers in about three and a half hours. Seven hours of reading time for 35 essays at the rate of $5.00 per hour equals an approximate reading cost of $1.00 per essay. Approximately ten percent of the essays are reviewed, adding another $.10 to the cost of each essay.)

Initial Use of the Competency-Based Scale

In the summer of 1975 the scale was tried out on a group of about 50 entrants. As a result, some changes in the scale itself were made, instructions to readers were developed, and a preliminary check was made on the consistency of ratings between the two readers for each essay.

In the fall of 1975 a full-scale trial was carried out. Essays were written by 241 entrants and read by seven teams of seven readers. (Each reader read two sets of essays.) An eighth reader served to review disagreements.

Results for this group of students (Table D2-1) show that while the whole scale was used, the majority of essays received ratings between 3 and 7. Readers considered few students to be writing publishable material; the majority were at the educated adult level. These students met the college's requirement for competency in writing. Readers judged that a substantial number of entrants needed to improve their writing but that only a few had serious problems. These results are not surprising. At entrance to college few students are skilled writers, and few should have serious problems in writing. Probably the proportions writing at an educated adult level and needing improvement vary from college to college, depending on selectivity in admissions. The table points to one feature of competency-based scales: not all points on the scale are or should be used. If all essays had indicated a need for improvement in writing, for example, then only scale points 3 and 4 would have been used.

Table D2-1

Ratings (Average of Two Readers) of Essays of Fall 1975 Entrants

N = 241

Competency Scale		N	%
Clearly publishable	9.0	0	0%
Publishable	8.0-8.9	2	1%
	7.0-7.9	8	3%
Educated adult	6.0-6.9	48	20%
	5.0-6.9	94	39%
Needs to improve	4.0-4.9	58	24%
	3.0-3.9	24	10%
Serious difficulties	2.0-2.9	6	2%
	1.0-1.9	1	0%

Average = 5.0   Standard deviation = 1.1

Consistency of Readers

To increase the consistency or reliability of ratings, the final rating assigned an essay is the average of the two readers' judgments; if the review reader participates, the final rating is the average of all three readings. Such a procedure insures that the student communicates with more than one other person and protects him or her against individual idiosyncracies of readers. (Statistically it is equivalent to doubling the length of the test.)

Table D2-2 deals with the agreement of the readers. The average ratings assigned by different readers indicate how they differ in severity of judgment. Readers 1, 2, 4, and 7 are typical; readers 3 and 5 are generous in their judgments while reader 6 and the review reader are severe. Readers also differ in nonsystematic fashions; most educated adults will have their own idiosyncracies in the weight they place on different errors and different excellencies. One reader may be so impressed by well thought-out organization that errors in grammar are regarded as trivial while another may be so annoyed by misspelled words that excellent style in writing goes unperceived. Table D2-2 indicates substantial but not perfect agreement among the readers. Three of the original readers agreed within one scale point (a difference of one writing level equals two scale points) with the other reader for three-quarters or more of the essays. The other four readers agree in just less than two-thirds of the cases. Serious differences (at least three scale points or more than one writing level) occurred about a tenth of the time. Every serious difference was reviewed, and for three-quarters of the papers the review reader agreed with one of the original readers.

Table D2-2

Consistency of Ratings: Agreement Between Readers

Reader	Number of essays read	Average rating	Agreement between readers: Agreement (0-1 points)	Mild disagree (2 points)	Strong disagree (3 or more pts.)
1	65	4.7	65%	24%	11%
2	69	4.9	62%	22%	16%
3	105	5.8	77%	15%	8%
4	69	5.1	74%	17%	9%
5	37	5.8	84%	11%	5%
6	69	4.2	61%	25%	14%
7	66	4.9	64%	25%	11%
Review*	26	4.0	81%	8%	12%

* Since the review reader read only papers on which the original readers had disagreed by at least 3 points, an agreement with one reader automatically meant a disagreement with the other. Disagreements for the review reader mean a lack of agreement with either of the previous readers. Agreements for the review reader mean agreement with one or the other of the previous readers.

Another way of looking at consistency of ratings is through correlations of ratings for pairs of readers. Table D2-3 shows that with one exception readers' ratings are moderately correlated. Since, however, idiosyncracies of communication were expected, the final rating for any paper was based on the ratings of both readers. Thus correlations between readers represent the relationship between the halves of the final rating. The consistency of the combined rating based on both readers can be estimated by correcting the obtained correlation with the Spearman-Brown formula. This, in effect, estimates the reliability of two raters rather than one rater. Except for one reader, the estimated consistency of the total ratings is reflected in coefficients of .59 to .70, which approach a satisfactory level of consistency.

An attempt has been made to increase the consistency of grading still further through the development of training materials. A series of essays has been selected to represent points 3 through 8 on the nine-point scale. In addition, a training series of eleven essays has been provided. Potential readers can study the scale and the essays representing scale points, rate the training essays, and compare their judgments with those already agreed upon.

Judgments of Writing by Course Instructors

Do educated adults, basing their ratings on one essay, judge writing similarly to college faculty who base their ratings on writing in their courses? Agreement would testify to the consistency and the validity of both the initial readers' and the course instructors' ratings.

The two samples of writing differ in many respects. The writing-sample essay has the advantage that everyone writes on the same topic at

Table D2-3

Consistency of Ratings: Correlations Between Readers

Readers	Number of essays read	Correlation between readers	Estimated correlation of final rating*
1 vs 2	34	.42	.59
2 vs 4	35	.10	.18
3 vs 4	34	.15	.26
3 vs 5	37	.53	.69
3 vs 6	34	.42	.59
6 vs 7	35	.45	.62
1 vs 7	31	.54	.70

* Since the rating assigned each student was based on ratings by both readers, the correlation between readers has been corrected by the Spearman-Brown formula for doubling the length of a test.

the same time; but such standardization is more apparent than real--for some students the topic may be unsuitable, others may be writing at a time when they are not prepared or under conditions that are not typical. Writing in a course has the disadvantage of lack of control over the conditions under which the students write, but it should have many advantages. The students may have chosen the course, the topics on which to write, and the conditions for writing. The sample of writing may be longer. College faculty should be practiced judges of student writing and know the standards of writing in their fields.

All instructors in the fall quarter of 1975-76 were asked if their courses involved writing (see Appendix 6j). Since the initial ratings of writing were based on two readers, the end-of-course ratings were to be based on a minimum of two courses involving writing. Of the 241 students who wrote essays, 167 (69 percent) were taking at least two courses involving writing. During the second half of the course the competency-based scale for judging writing was sent to the course instructors of these students. For 88 students (53 percent) at least two ratings were received; 60 others (36 percent) received one rating while 19 (11 percent) received none. The study of agreement between initial readers and course instructors is based on the 88 students who received at least two ratings. Students with two ratings differ slightly from the students for whom two ratings are not available: more of the former have high (over 5) or low (under 4) ratings while more of the latter have average (4 and 5) ratings.

Advising must be taken into account in comparing judgments of initial readers and course instructors. At the beginning of the quarter all students and their advisers were notified of the student's level of competency in writing. Students who were notified that they had met the college's standards for competency in writing could assume that they were performing satisfactorily. Presumably the notice had little effect on them.

On the other hand, students who were told they needed to write better not only were likely to become more aware and to take more care but also were so counseled by their advisers. A questionnaire (Appendix 6k) sent to faculty advisers was answered by 11 of the 20 advisers who had more than one student needing work on writing. In every case the adviser discussed writing with the student and made suggestions for improving skills. Of the students needing work in writing, 61 percent took at least two courses requiring writing; in addition, they took courses related to work in writing: 52 percent took freshman core courses, 11 percent took literature courses, and 11 percent took a workshop in reading. Thus, it appears that the advisers did discuss writing with those advisees who needed work, and they recommended particular courses. The majority of students who needed work in writing planned to include it in their course schedules.

Table D2-4 presents the ratings of initial readers and course instructors for the total group, for those initially rated below 5.0, and for those initially rated 5.0 or above. The average initial rating was 5.1, compared with 5.0 for all students, and 65 percent were initially rated competent, compared with 63 percent in the total group with a very similar distribution of ratings. This distribution of ratings indicates that most students were not limited by having been initially rated so high or so low that change upward or downward was impossible.

Table D2-4

Distribution of Initial and Course Ratings for Students Rated By at Least Two Initial Readers and Two Course Instructors

Ratings	Total group		Below 5		5 and above	
	Initial	Course	Initial	Course	Initial	Course
8-9.0	1	1		0	1	1
7-7.9	4	13		2	4	11
6-6.9	23	36		11	23	25
5-5.9	29	19		5	29	14
--						
4-4.9	17	12	17	6		6
3-3.9	10	7	10	7		0
2-2.9	4	0	4	0		0
1-1.9	0	0	0	0		0
Average	5.1	5.6	3.8	5.2	5.8	6.0
N	88		31		57	
t-value	4.7**		6.7**		1.5	

** $p < .001$

For the total group the average course rating is 5.6, a gain of 0.5 scale points over the initial rating. When the total group is split into those who met the competency level initially and those who did not, a different picture emerges. Those initially rated competent in writing (5.0+) had similar initial and course ratings while those not so rated (below 5.0) had higher course than initial ratings. Of those initially rated competent, 90 percent were also rated competent by course instructors; on the other hand, 58 percent of those who had previously been rated as needing improvement were now rated as competent. (The t-values listed in the final row will be discussed below.)

These results can be studied from two viewpoints. First, the group initially rated competent can be studied for consistency of ratings between initial readers and course instructors, since motivation for writing was not changed. Second, the group initially rated below 5.0 can be studied for the effects of motivation, advising, and relevant work on effecting changes in writing over a two- to three-month period.

In Table D2-5 initial readers and course instructors are more in agreement than initial readers (69 percent agreement, 20 percent mild disagreement, 11 percent strong disagreement) in that strong disagreement has been absorbed into the mild disagreement category. With initial readers, however, agreements are between two individuals; when initial readers and course instructors are compared, each category is represented by the average of two individual judgments, which should be more reliable. Nevertheless, considering the differences in motivation and sampling in the two situations (diagnostic test and course work), the amount of agreement is impressive. Not only does the amount of agreement

Table D2-5

Consistency of Ratings: Agreement Between Initial Reader and Course Instructor* for 57 Students Originally Rated 5.0 or Above

	N	%
Agreement (0-1.0 points)	39	68%
Mild disagreement (1.1-2.0 points)	18	32%
Strong disagreement (2.1 or more points)	0	0%

* Average of two or more readers or instructors

between readers and course instructors indicate that they are judging competency in writing similarly, but it also implies that the competency-based scale can be used not only for short samples of writing but also for long papers and for quality of writing over a period of time.

In five courses the instructor rated at least ten students whose writing had initially been judged competent. In Table D2-6 the average of two initial readers is compared with each instructor's ratings, more closely approximating the original comparison of two readers with each other. For the instructors as a group, agreement is as good as that for initial readers. In three courses, instructors agreed with the initial readers. In one course, the instructor rated the writing higher than did the initial readers; in another the instructor was a more severe judge. Disagreement was thus produced by systematic differences in severity of judgment. The training materials may be useful for the faculty in developing comparable standards.

Gains from Ratings by Initial Readers to Ratings by Course Instructors

Table D2-7 clearly makes the point that instructors rated the work of students whose writing was originally judged to need improvement higher at the end of the first quarter's courses. Almost two-thirds received higher ratings while only one student was rated lower. This excess of higher over lower course ratings as compared with initial judgments is statistically significant. A statistical test known as the dependent t-test was performed on each of these three groups: the total group, those initially below 5.0, and those who scored 5.0 or above. The dependent t-test determined whether the course ratings were significantly higher than the initial ratings. The resultant t-values are presented as the bottom row in Table D2-4. Both the

Table D2-6

Consistency of Rating: Agreement Between Initial Readers (Average) and Instructors In 5 Courses for Students Initially Rated Competent (5 and Above)

Course	N	Initial Average	Final Average	Agreement (0-1.0 pts.)	Mild Disagree. (1.1-2.0 pts.)	Strong Disagree. (2.1 + pts.)
Biology	14	5.6	5.7	86%	0%	14%
Communication	10	6.2	6.0	100%	0%	0%
Education	10	5.8	7.1	60%	10%	30%
Literature	11	5.6	5.1	82%	18%	0%
Psychology	19	5.9	4.4	37%	32%	32%
Total	64	5.8	5.5	69%	14%	17%

scores of the total group and those initially below 5.0 rose significantly. The change in scores of those initially at or above 5.0 was not statistically significant. There are three qualifications to these analyses. A phenomenon known as "regression toward the mean" explains the ever present finding that individuals initially below the mean tend to increase their scores and those initially above the mean tend to decrease. Furthermore, both "ceiling" and "floor" problems are encountered here. For instance, the student who makes 8.0 on the initial test cannot increase much due solely to the constraints of the scale; this is called a "ceiling" effect. Likewise, a student scoring 2.0 on the initial testing cannot decrease his or her score appreciably. Finally, since there is no control group present here (students who do not take intervening instruction), one may only hypothesize that the change in student writing is caused by the remedial instruction.

An initial diagnosis that writing needed to improve apparently led to advising of students and to students' planning work with the result that 65 percent improved in writing (and only one student lost ground) and that 58 percent were judged competent in writing by course instructors. The competency-based scale not only initially identified competency in writing as a requirement for graduation but also provided a useful diagnostic instrument for faculty advisers and students.

The story for those originally rated competent is different; here the excess of higher over lower ratings (13 to 5) is not significant. What is meaningful is that if random errors in grading accounted for the differences between initial readers and course instructors, the group with high initial ratings could have been expected to receive lower ratings from instructors

Table D2-7

Distribution of Differences Between Average Ratings of Two or More Readers of Initial Essays and Average Ratings of Two or More Instructors at End of Courses

	Total Group		Initial rating below 5		Initial rating 5 and above	
	88		31		57	
Gains (discrepancy of 1.5 or more points in favor of end-of-course rating)						
3.5 to 4.0	1	38%	1	65%	0	23%
2.5 to 3.0	5		5		0	
1.5 to 2.0	27		14		13	
Agreement						
0.5 to 1.0	20	56%	4	32%	16	68%
0.0	13		5		8	
-0.5 to -1.0	16		1		15	
Loss (discrepancy of 1.5 or more points in favor of initial rating)						
-1.5 to -2.0	6	7%	1	3%	5	9%
X^2, df=1	18.69*		17.19*		3.56	

*p<.01

and the group with low initial ratings to receive higher ratings. Since both groups received higher ratings from course instructors, this statistical regression to the average rating probably does not account for all the differences observed. The fact that course instructors rated the competent group (5 and above) slightly higher can be explained by some training or by better motivation in courses than on diagnostic tests or by better sampling of writing in courses.

Summary and Conclusions

A scale for judging competency in writing has been developed with the standard for competency set at the level of the educated adult. The scale can be used consistently by non-faculty readers and faculty on a variety of materials and in a variety of circumstances. Further generalizations of the use of the scale are under way; the scale and the training materials have been used in judging the competency in writing of adult Mexican-American applicants to an undergraduate program and a master's program in education. In the master's program the two readers agreed in their ratings of all of a sample of 20 applicants. The scale can also be used for evaluating writing for credit for learning based on prior experience. At Antioch a rating at the educated adult level (5 to 7) warrants a judgment of college-level competency or credit equivalent to a college course in writing. If credit for publishable expository writing is sought, the ratings should be at least 7. The scale is probably not suited for judging imaginative writing.

Two precautions should always be taken to reduce systematic differences in severity of judgment and individual idiosyncracies. First, two persons should read each sample of writing. (As we have seen, this is not an expensive procedure.) Second, readers should be provided with training materials consisting of materials exemplifying the scale points and samples of writing (already rated) on which to practice.

A final consideration is that readers should only rate the samples of writing; they should not determine the use made of their judgments. The institution should decide for itself what its standards are: What scale point represents for them the competency in writing of an educated adult? How many credits will be awarded for what level of performance?

STUDY D3:

THE USE OF EXPERT JUDGMENT IN THE PRE-GRADUATION CONFERENCE

Paul I. Jacobs, Thomas A. Edison College

Introduction

The Pre-Graduation Conference at Thomas A. Edison College is a face-to-face discussion, lasting approximately one hour, between each candidate for the B.A. degree and two college faculty members. One faculty member is selected to discuss with the student certain topics in his or her "area of concentration" (Humanities, Social Sciences, or Natural Sciences/Mathematics). The second faculty member covers more briefly topics in one of the other two areas.

The essential purpose of the Conference is to ensure that every degree candidate demonstrates an acceptable grasp of basic concepts both in the Area of Concentration and in the second area he or she selects.

As an outcome of the Conference, each faculty member arrives at a judgment of "pass" or "defer with recommendations" in the area for which he or she has been given responsibility. "Pass" means the faculty member believes the student has an acceptable grasp of basic concepts in the area under consideration. "Defer with recommendations" means the faculty member does not believe the student has an acceptable grasp of basic concepts. This judgment, when given, is accompanied by specific suggestions for further readings, courses, etc., intended to bring the student to the "pass" level. In addition, each faculty member briefly records some of the considerations that led to this judgment. A student may in this way receive a "pass" in both areas, just one area, or neither area.

The Pre-Graduation Conference Program was started in April 1975. In February 1976, changes were made in the information sheets about the Conference that are given to students and faculty members, and in the student's application form for the Conference. These changes were intended to convey more clearly than before, the basic purpose of the Conference and to provide the student with more guidance in his or her choice of topics for the Conference. Copies of the newer information sheets and application form are reproduced in Appendices 61, 6m, and 6n. A number of students made their applications on the older forms, which they received prior to February, although their actual Conferences may not have taken place until May.

The judgments faculty members make in a Pre-Graduation Conference involve a large subjective element that can be reduced but not eliminated by further training and by providing them with more adequate guidelines. The purpose of this research project was to determine under the present conditions the agreement among independent judges within the Pre-Graduation Conference setting. More specifically, the research question was whether a second faculty member in a given area, unaware of the judgment of a first faculty member, would agree with that judgment.

The Research Procedure

It was decided to collect the research data under circumstances that depart as little as possible from the ordinary routine of the Pre-Graduation Conference. Each student, therefore, was not subjected to separate Pre-Graduation Conferences run by different faculty interviewers. Instead, faculty participants made their respective judgments in an actual Conference, then other faculty members were asked to make judgments independently after listening to a tape recording of the Conference.

The research study was directed at the population of all 38 Conferences carried out between January 1 and May 25, 1976. The first column of Table D3-1 shows how often each area (Social Sciences, Humanities, and Natural Sciences/Mathematics) occurred in this population of Conferences. Proportionate samples were randomly drawn for each area separately, as shown in the second column of Table D3-1, with a total sample of 60. In this procedure both, one, or neither of the two areas of a given student's Conference might be included in the sample.

In the actual administration of the Pre-Graduation Conference Program, the same faculty assessors are frequently called upon to serve for more than one Conference. Thus, for example, as the third column of Table D3-1 shows, only 20 different assessors were used for the 35 Conferences that involved Social Sciences. For this research study, it was decided to use as tape listeners those assessors who had served most frequently in the actual Conferences. It was further decided that each tape listener would judge four different Conferences in which he/she had not taken part, resulting in the number of listeners shown in the last column in Table D3-1.

In the actual administration of the Pre-Graduation Conference Program, a faculty assessor is chosen with regard to the particular topics a student lists in the application for the Conference. Thus, if a student whose area of concentration is Social Sciences lists Nineteenth Century European History as a topic, the student would be matched with an assessor who teachers History, preferably one particularly familiar with Ninetheenth Century European History. In this research study, however, a faculty member who teaches any Social Science, such as Psychology, might be matched as a tape listener with a

Table D3-1

Number of Conferences and Assessors in Three Academic Areas

	Number of Instances in Population of Conferences	Number of Instances in Sample of Conferences Drawn for Study	Number of Different Assessors in Population of Conferences	Number of Different Assessors Used as Tape Listeners in Study
Social Sciences	35	28	20	7
Humanities	25	20	13	5
Natural Sciences/ Mathematics	16	12	10	3
Total	76	60	43	15

student listing a topic in History. This matching procedure was felt to be in accord with the intent of the Conference to assess whether a student has an acceptable grasp of basic concepts in an area as broad as Social Sciences. Furthermore, this procedure was intended to provide a more stringent test of the agreement among independent judges.

Similarly, in the case of a Conference in the Humanities, the tape listener in this study would be an instructor in the Humanities but not necessarily matched with regard to the subject matter of interest to the student. The area of Natural Sciences/Mathematics was subdivided into Life Sciences (e.g., Nutrition, Environmental Health) and Physical Sciences/Mathematics. Within each of these subareas, tape listeners were freely matched with students.

Each tape listener heard the part of the recording dealing with the area about which he/she was concerned, made a judgment of "pass" or "defer," and then gave some indication of the basis for that judgment on the same kind of form used in actual Conferences. The tape listener did not hear any evaluative comments of the faculty judge in the actual Conference.

In three cases, the randomly selected tapes were of unacceptable audio quality and were replaced by other tapes selected at random. In 14 cases, the intended judgments were not carried out for various administrative reasons, chiefly that of the scheduled tape listener not keeping an appointment. The final sample, therefore, involved 46 rather than 60 instances of independent judgments that could be compared, and involved 12 rather than 15 tape listeners.

Results

The overall "pass" rate in the entire population of Conferences under study was 76 percent. This is identical to the overall pass rate for the actual Conferences in the sample of 46 cases for which independent judgments are available. This strongly suggests that the sample, even with loss of some cases, is representative of the sample from which it was drawn.

Table D3-2 shows how the actual Conference judgments in the 46 cases compares with the tape listeners' judgments. In 34 of the 46 cases or 74 percent, there was agreement. This degree of agreement is perhaps not as high as one might hope, but may be considered a lower bound of the degree of agreement that might have been obtained had tape listeners been matched with students in terms of specific topics.

How adequate is 74 percent agreement? A statistical answer can be approached through a chi-square analysis, which reveals 74 percent to be significantly higher than the 59 percent agreement that would be expected on the basis of chance alone (chi-square with Yates' correction = 4.61, $p < .05$). Such an analysis would assume independent frequencies in each cell, a condition that is not strictly met, since the same listeners judged more than one tape. Additionally the phi coefficient is .37, indicating the relationship to be only moderate although positive and statistically significant.

Aside from statistical considerations, however, how meaningful is the 74 percent agreement in terms of the operation of the Pre-Graduation Conference Program? The data imply that in about one case in four a student who would be passed by one judge would be deferred had another one been chosen, or vice versa. While greater agreement would certainly be desirable, and should be striven for, the level of agreement already achieved in this innovative program may be viewed as an accomplishment. By way of comparison, a study published in the New England Journal of Medicine found that "...when the opinion of a second surgeon was mandatory prior to an elective, non-emergency operation, 17.6 percent of the operations were not recommended by the consulting surgeon." (The New York Times, February 6, 1976). This may be translated into an agreement rate of only 82.4 percent in a situation where the consequences of failure to operate when necessary or of operating when unnecessary might be assumed to be more drastic than the consequence of an improper decision on a Pre-Graduation Conference.

Another point to keep in mind is that the situation for the actual assessor differed in a number of ways from that of the tape listener, and that these differences would tend to reduce the amount of agreement found. The actual assessor could take into account the student's verbal and nonverbal behavior during the entire Conference covering two areas, ask what questions he or she wanted to, know that the decision he or she made would directly affect the student, and know that a decision to defer must be personally conveyed to the student in a face-to-face meeting. The tape listener, on the other hand, was exposed only to the student's verbal behavior during the discussion of just one area, could only hear the answers to someone else's questions, know that the decision he or she made was for research purposes only, and did not need to personally convey to the student a decision to defer. As an indication of the importance of the differences between these two situations, it may be noted that while the pass rate for

Table D3-2

A Comparison of Actual Conference Judgments With Tape Listeners' Judgments

Actual Conference	Tape Listener	Frequency
Pass	Pass	27
Pass	Defer	8
Defer	Pass	4
Defer	Defer	7

the actual Conference in the sample was 76 percent, the pass rate for the tape listeners was only 67 percent.

It is also of interest to determine the degree of agreement between two independent judges within each of the three areas separately. The relevant data are presented in Table D3-3.

While overall there is 74 percent agreement, within the Humanities area agreement is as high as 87 percent, and within the Social Sciences area it is as low as 65 percent. These apparent differences among the three areas should not be considered definitive, given the small numbers of cases in each area and the even smaller numbers of tape listeners. A review by this writer of the reasons given for their judgments by both actual Conference assessors and tape listeners suggests a tendency for judges in the Humanities to be more concerned with a student's grasp of general principles, and for judges in the Social Sciences and in Natural Sciences/Mathematics to be more concerned with a student's grasp of specifics. Here are some non-randomly chosen examples:

> "Was unable to discuss any cultural aspect in detail. Nor was she ready to discuss music, art, theatre, or literature on a broad basis."
>
> -- Humanities

> "She could not describe a nerve cell, a typical cell, how nerve messages were transferred, or why a stroke injury did not heal."
>
> -- Natural Sciences/ Mathematics

> "The student was unclear on positive reinforcement practices, e.g., could not explicitly define continuous reinforcement, intermittent reinforcement, or when such practicies were used."
>
> -- Social Sciences

Table D3-3

A Comparison of Judgments by Area

Actual Conference	Tape Listener	Frequency		
		Humanities	Social Sciences	Natural Sciences Mathematics
Pass	Pass	10	12	5
Pass	Defer	1	5	2
Defer	Pass	1	2	1
Defer	Defer	3	1	3
Pass Rate, Actual Conferences		73%	85%	64%
Pass Rate, Tape Listeners		73%	70%	55%
Percent Agreement		87%	65%	73%
Chance Agreement Level		60%	65%	51%
Chi-Square (corrected)		3.50	.30	.74
Phi Coefficient		.66	.03	.24

A more formal content analysis of the reasons expert judges in each area give for their judgments would be of interest, but is beyond the scope of the present research project.

An analysis was also made of agreement between judges when the students are separated into groups of those who applied using the old and those using the new information sheets and forms. Four cases of students retaking the Conference were excluded from this analysis. The results are presented in Table D3-4.

The change from old to new procedures, intended to more clearly convey the basic purpose of the Conference and to provide the student with more guidance in choice of topics, has produced a substantially higher pass rate. Paradoxically, the agreement between actual assessor and tape listener has fallen substantially under the new procedure. The resolution to this paradox would seem to be the recognition of the large cluster of students who were passed in the actual Conference but deferred by the tape listeners. This circumstance, presumably due to the previously discussed difference between the actual Conference situation and the tape listening situation, placed an upper limit of 75 percent on the degree of agreement that could have possibly been found using the new procedures. The implication of this line of thought is that higher agreement with the new procedure would have been obtained if the two independent judges in a given area were both physically present and both involved in the actual Conference.

Summary

The Pre-Graduation Conference is designed to allow candidates for the B.A. degree at Thomas A. Edison College to demonstrate an acceptable grasp of

Table D3-4

A Comparison of Judgments in "Old" and "New" Application Procedures

Actual Conference	Tape Listener	Frequency	
		Old	New
Pass	Pass	13	12
Pass	Defer	0	6
Defer	Pass	3	1
Defer	Defer	6	1
Pass Rate, Actual Conferences		59	90
Pass Rate, Tape Listeners		73	65
Percent Agreement		86	65
Chance Agreement Level		54	62
Chi-Square (corrected)		8.79**	.74
Phi Coefficient		.74**	.10

** p < .01, 1 degree of freedom

basic concepts in two of the following three areas: Humanities, Social Sciences, and Natural Sciences/Mathematics. In the Conference one college faculty member in each of the two areas holds a discussion with the student, culminating in a judgment of "pass" or "defer." In this research project other faculty members listened to tape recordings of Conferences, then made independent judgments of "pass" or "defer." In 74 percent of 46 cases, the tape listener's judgment was the same as that of the faculty member who participated in the actual Conference. Agreement was greater within the Humanities area than in the other areas. Some implications of the results are discussed.

Chapter VIII

CONCLUSIONS AND IMPLICATIONS

Warren W. Willingham

The results of the validation stage of the CAEL project have been summarized in the relevant chapters of this report. This final chapter attempts to focus upon the main conclusions that might be drawn and the implications of the findings with respect to educational practices and further need for research and development. Because this validation work has involved a major effort on the part of a number of institutions, any summary of the conclusions and implications runs some danger of appearing to overgeneralize from this particular set of studies. Validating the assessment of experiential learning is, like all other educational assessment, a problem so broad as to engage all higher institutions on a continuous basis. Useful though this validation work may be, it is important that it be interpreted in an appropriate context considering the purpose and scope of the studies and taking into account important qualifications that may limit generality and condition conclusions.

Chapter I discussed the general purpose of this validation stage of the CAEL project. Briefly, it was to try out experimentally in a number of colleges assessment materials that had been developed in the previous year and to carry out a series of field studies that would help in the evaluation of the various assessment methods suggested. Chapter III considered in some detail the meaning of "validation" when the great diversity of experiential learning permits no standard instrument or uniform technique. For that reason this validation work is concerned with the appropriateness of materials in diverse educational settings and with the general soundness of principles and practices.

A major CAEL objective from the outset has been to aid in the development of assessment methods that would serve the students' educational needs but also protect the student's interests by supporting strong standards in sound programs. Throughout the project there has been heavy emphasis on quality assurance and improvement, not merely developing and disseminating materials and procedures. So it has been a conscious intention that this validation work should help to instruct current practice and guide further study and improvement. With that in mind, this final chapter addresses three questions:

-- What are the strengths and weaknesses of the three particular types of assessment examined here (interpersonal skill, prior learning, and work competence)?

-- What other implications can be drawn from the findings; i.e., implications concerning particular aspects of assessment practice or important problems that cut across many applications?

-- What summative impressions seem warranted concerning the general usefulness of the research and development model described here?

These three questions are addressed in successive sections of this final chapter, but it is useful first to mention briefly some important qualifications. These were discussed at the end of Chapter III; they are worth recalling. The most important qualification can be stated as follows. In attempting to collect common information that may have some useful generality, field research is almost inevitably nonrepresentative in some degree. It is not clear that the faculty and students involved in this work are completely representative of all faculty and students who might be interested in such programs. It is not clear that the participating institutions are wholly representative of those nationally who are now or might subsequently become involved in such experiential learning programs. Finally, research procedures and materials do not exactly match those likely to be encountered in a particular real situation. Typically, they are probably less complete and detailed because they focus on particular instances or aspects of a procedure under study.

One important respect in which some of the assessment procedures here did not match a routine institutional procedure was the lack of direct student involvement in several of the assessment studies. In Study C2, for example, inadequate documentation may be a source of unreliability, and in a real situation an assessor might well have insisted upon more information (see p. VI-15).

Because of these considerations, results of any research or evaluation will not necessarily apply to a particular institution, nor can specific aspects of findings necessarily be taken at full face value. It is important to remember, however, that it is the gross character of the research situation and the findings that matter. Researchers report results that they believe to be reasonable, but in the last analysis it is the reader who must decide to what extent conclusions may need to be tempered.

Also, it is important to remember that these results should be compared to results typical of human judgment, not standardized examinations. It is generally appreciated that faculty often disagree; previous research cited here has documented the extent to which classroom grades are unreliable. Where the findings of this study show lack of consistency there is little basis for assuming that such inconsistency is found only in nontraditional programs.

Assessment Strengths and Weaknesses

Assessing Interpersonal Skills

There were three studies concerning the assessment of interpersonal competence. One (A1) involved judging the extent to which instances of acquired interpersonal competence (briefly described in written documents) were creditable in different degree programs. Another (A2) involved assessing interpersonal competence in a counseling situation by two methods: an evaluation interview and a role-playing simulation. A third study examined the assumption that assessment has a beneficial effect on student self-awareness.

Without much doubt, the clearest strength in the assessment procedures evaluated in these three studies was the ability of independent faculty judges to agree fairly well on the degree to which students exhibited interpersonal competence (in this instance, in a counseling situation). This resul certainly suggests that these faculty knew the competence when they saw it. Furthermore, the results indicate that this competence can be assessed quite adequately either through direct observation of simulated performance or through an evaluation interview. This finding is potentially important because an interview procedure might be significantly more flexible and less costly in assessing a variety of interpersonal competences than would be especially designed role-playing simulations or observation of actual performance in real situations.

The data suggest an important weakness in the assessment of interpersonal competence; that is, that skill of this kind is not clearly delineated in the minds of the faculty. There are several indications. First, in Study A1 there was very poor faculty agreement as to whether particular instances of interpersonal competence were creditable in particular programs. Furthermore, judgment varied quite widely across departments and institutions as to how much credit particular instances of acquired competence might be worth. Second, in Study A2 faculty were not able to distinguish different aspects of interpersonal competence in a counseling situation. And in Study A3 there was no evidence that assessment of interpersonal competence improved the student's ability to describe his or her competence and its implications. This result suggests either that the competences could not be delineated or the students could not relate the delineated skills to personal goals.

Taken together these results suggest that interpersonal competence can be reliable assessed, but there is likely much work to do in clarifying what the competences are and how they relate to educational objectives and degree programs. CAEL assessment reports in this area contribute to a useful conceptual framework concerning the role of interpersonal competence in experiential learning. Future work needs to focus especially upon empirical research that will help to simplify the measurement of this obviously complex construct. An important issue, for example, is the extent to which different types of interpersonal competence (counseling skill, leadership, persuasion) can be reliably distinguished through the types of evaluation interviews or observational techniques employed here.

Assessing Prior Learning

There were four studies concerning the general strategy of using the portfolio in assessing prior learning. The first study broached the fundamental question as to whether reasonably similar credit evaluations result when a portfolio is judged in two independent evaluations. Two studies examined how well faculty agree on what is creditable and how prior learning ought to be documented. The final study examined the possible educative benefit of developing a portfolio; i.e., whether students who have gone through the assessment process seem to exhibit a greater degree of self-awareness than students who have not done so.

Results of two of these studies represent clearly positive strengths in the assessment of prior learning. First, there was quite respectable agreement in the credit awards based upon successive independent evaluations of portfolios. This finding is important because it indicates that, by and large, students are not subject to credit awards that vary substantially depending upon whoever happens to evaluate the portfolio. By the same token, this finding suggests that these institutions have taken reasonable steps to see that faculty evaluating portfolios do so within a generally comparable frame of reference. This is not to say that the reliability of assessment of experiential learning cannot be improved. It can be and should be, but this is very true of higher education generally.

A second important strength in portfolio assessment is suggested by the positive finding concerning the educative benefit of this assessment process. In most institutions students who had participated in portfolio assessment did demonstrate in the eyes of objective judges a greater degree of self-awareness regarding educational objectives and personal plans. This result reinforces a basic tenet of the CAEL project--that assessment can and should have direct personal benefit beyond simply certifying learning.

The main weaknesses in portfolio assessment suggested by this field research concern the assumptions underlying the assessment process. There was only modest agreement among faculty within institutions as to what constitutes college-level learning. There was often disagreement as to what documentation should be required for learning and evidently no clear understanding as to what constitutes authentic documentation.

The consistency in credit awards and educative value of portfolio assessment are critical findings, but the results do suggest a need for a sound basis for faculty consensus if assessment practices are to be generalized to larger groups of students and faculty. The weaknesses cited above can possibly be attributed partly to the fact that no one can be an expert concerning each of the types of learning represented in Studies B2 and B3. But this begs the question as to who speaks in a common voice on such a fundamental institutional issue as the types of learning that can be credited and what constitutes adequate evidence.

Or these weaknesses may have resulted from a use of a group of faculty in Study B2 and Study B3 who were simply somewhat less experienced than was the case in Study B1. In any event, narrative information provided by the colleges suggest that institutional policies and procedures and staff experience are still in a developing stage in many institutions. Data reported here on such a critical question as, "What is creditable?" seem consistent with that assumption. One implication is a need for better institutional guidelines regarding assessment practices; another is a need for faculty development programs--a need that seems likely to continue for some time.

Assessing Work Competence

Studies in this area drew upon a specialized assessment model designed to identify fairly specific competences related to particular occupational programs. One study concerned the extent to which work supervisors agreed with student claims of learning acquired. Another study examined the faculty's ability to relate work competences to curriculum areas. A third study examined faculty agreement in awarding credit for such competences.

A strength of this assessment model is the fairly good agreement found between students and supervisors as to what the student has actually learned. That is, there is little evidence of unrealistic claims by students. Another strength was reasonably good consistency among raters in discriminating the relative level of competence among different students. A fairly important weakness, however, was the fact that raters differed substantially in the general level of credit they characteristically awarded (a result seemingly similar to that of Study B2). There was also limited evidence of faculty agreement as to the relevance of specific competences to particular curriculum areas. Both of these weaknesses may have resulted from limited familiarity of most of these particular faculty participants with the assessment of experiential learning.

The results of the field research in this area suggest that the detailed checklist procedure for identifying and crediting specific work competences related to particular occupational programs is not something that can be carried out easily with little experience. It appears that the approach is reasonably sound for the limited purpose intended, provided faculty have a consensual framework that specifies the general level of credit that is appropriate and how credit for such learning fits the curriculum. Again, there is indication of the same need for institutional guidelines and faculty development that was cited above. These results suggest that there is danger of fairly substantial discrepancies in credit awards if faculty judges are inexperienced.

Other Findings and Implications

In addition to the aforementioned findings that provide some evaluation of the three main types of assessment examined, there were other results especially evident in particular studies that have important implications for assessment generally. Six such results are described in the following paragraphs.

Significant Differences Among Raters

As discussed in Chapter III, consistent differences among faculty members can be an important source of error in expert judgment. In two studies reported here (C1 and C2) the same pair of raters judged each example of learning so it was possible to examine systematic differences between raters. The principle applies to all judging situations. In these studies the rater-error was quite pronounced.

All judges, but particularly inexperienced judges, may make radically different credit awards because of important differences in the assumptions they make. For example, different raters may make different assumptions regarding what is creditable, whether a particular instance of learning overlaps with other credit, whether learning is still current, and so on. When two raters rate independently, such different assumptions become apparent. Discrepancies can be rectified in discussion if there are adequate policies and guidelines to call upon. The important implication is that two opinions should be sought wherever possible, especially if raters are relatively inexperienced. Furthermore, it is better if the two opinions are reached quite independent of one another because independent judgments are much more likely to smoke out erroneous assumptions in assessment than if one rater simply reviews the judgment of another. Independent ratings are also a critical issue because of the cost involved. A reasonable tactic may be to obtain independent assessments in the development phase of a program and thereafter use two independent assessments only for periodic quality control.

Significant Differences Among Institutions

In several of these studies of faculty agreement there were quite substantial differences among institutions. Study B1 provides a good example. In that evaluation of portfolio assessment, faculty within individual institutions agreed fairly well on how much credit particular portfolios were worth, but there were quite large differences across institutions in the average credit awarded for the same set of portfolios. Of course, in a pluralistic system of higher education there is nothing inherently inappropriate in different institutions being willing to award different amounts of credit for the same learning. On the other hand, wide differences in credit practices from one institution to another do beg the question as to whether standards are receiving the attention they deserve and whether each institution has a satisfactory rationale for its credit policy. It would seem quite proper for such a rationale to receive careful attention in institutional guidelines regarding credit policies for experiential learning.

Why Documentation?

Study B3 was the only one of this series specifically concerned with documentation, but in its broad interpretation, documentation applies to many types of assessment. Informal communication with participating institutions suggests that many faculty are not especially interested in documentation. Some profess to concentrate on measuring learning, not documenting experience. Other institutions may emphasize contracting for further development of competences rather than crediting prior learning and maintain that documentation is therefore not judged to be important for record purposes.

Institutional philosophies and practices vary greatly but these arguments do not take fully into account different functions that documentation can serve. For example, documentation can refer to an assembling of information in which case documenting may be best identified as a learning experience. In some instances documentation (e.g., an evaluative letter from a work supervisor) is synonymous with evidence for or assessment of learning. Finally, documentation can be perceived primarily as a record or paper trail of a process. These various functions or purposes of documentation may serve the student or the institution and they are to some extent independently useful. The data and informal information collected in connection with Study B3 do not suggest that most institutions have made an adequate effort to clarify what forms of documentation would be useful to whom for what purpose.

The Halo Problem

Several lines of evidence suggest that in some instances faculty raters participating in this field research found it difficult to make consistent discriminations of different aspects of an overall judgment. For example, in Study A2 faculty were not able to make consistent discriminations among students with respect to the seven aspects of counseling skill that were rated. This was true both in the case of the interview method of assessment and the role-playing assessment. Such findings are quite common in the extensive literature of human judgment, personnel ratings, etc.

The fact that raters often cannot really discriminate different aspects or traits of an individual or object being rated is often called the "halo effect." That is, the rater makes one overall judgment and any additional ratings of supposedly different aspects are merely reflections of the overall judgment. Thus these results of Study A2 are quite consistent with the frequent finding that judges simply are not able to make fine distinctions about the same object or person. The main implication is that faculty assessing experiential learning should be skeptical about the reliability of any multiple ratings of different aspects of one document or performance. It is especially hazardous to allow credit distinctions to ride on apparently real but probably unreliable differences among several judgments that are really part of the same halo. It may be useful to rate different aspects of a learning example, but unless there is evidence to the contrary it is safer to assume that the overall score is the reliable judgment that should be recorded.

Educational Value of Assessment?

Two studies reported here give conflicting evidence regarding the educational value of the assessment process. In the study of the portfolio process (B4) those students who had gone through assessment produced questionnaire responses that were judged to reflect significantly greater self-awareness than was true of the responses of students who had not gone through assessment. No such effect was found in Study A3 that involved assessment of interpersonal skill. The discrepant results between these two studies is of some interest because they were in most respects quite similar studies except for the type of assessment process involved. It seems likely that the portfolio assessment yielded a positive effect because it is specifically formulated to help the student understand personal goals and educational objectives. Even at that, the positive effect of assessment was not clearly differentiated across questions to which the student responded or across educational programs. This type of experiment was a demanding test of the beneficial effects of assessment and it is encouraging that a difference between the pre- and post-assessment group was clearly significant in one study at least, but the findings indicate that this effect was not large. More sensitive methods for evaluating the beneficial effects of assessment would be extremely useful. These methods in turn might help to focus assessment practices more effectively upon outcomes of value to the student.

What Choice of Method?

Study A1 demonstrated fairly clearly that faculty have considerable difficulty agreeing on the creditability of specific instances of interpersonal competence when the competence must be assessed on the basis of verbal documentation of the sort usually found in a student portfolio. In Study A2, on the other hand, judges agreed reasonably well when they assessed interpersonal competence in a face-to-face relationship with the student. A specific implication is that some competences may simply require direct interaction with the student if a sound assessment is to result. The more general implication is that assessment may give quite different results depending upon the method of assessment used. Furthermore, there are likely to be important cost implications that require close analysis to evaluate accurately. For example, the following four approaches to assessment have quite different administrative and measurement characteristics: observation of performance, self-report, oral examination, and testimony.

It will be recalled from Study A2 that two methods of assessing interpersonal competence gave quite similar results, It may be, however, that both role-playing and the evaluation interview are basically similar performance tests in which the assessor's judgment is heavily influenced by the direct observation of the student's behavior in the face-to-face situation. Study A1, on the other hand, used written documents that usually involved a combination of self-report and testimony. The results appeared to be much more erratic than was true of Study A2.

In this particular instance the apparent difference in the reliability of assessment probably results from a subtle but important difference in the type of question posed by the two rating situations. In Study A2 the level of interpersonal competence was indicated on a rating scale. In Study A1 the level of interpersonal competence was indicated by the amount of credit the competence was worth. The latter type of assessment was probably confounded by some judges rating level of work experience in addition to interpersonal competence itself. The fact that these two studies can yield such different results underscores the importance of studying the characteristics of different methods of assessment, their strengths and weaknesses, their direct and indirect costs, and what methods are most appropriate for what assessment situations.

Reflections On the Validation Model and Future Work

As explained in Chapter I of this report, the CAEL project initially faced the difficult task of working toward early assistance to rapidly developing programs involving the assessment of experiential learning. The task was doubly difficult because of two partly incompatible goals: 1) developing sound and useful assessment methods in a way that would respect the diversity of experiential learning and not standardize assessment methods and 2) fostering quality assurance in assessment practices so as to support the development of sound programs of experiential learning and engender confidence in those programs.

In working toward these objectives CAEL devised a unique developmental model. This model placed priority on widespread participation in order to avoid a narrow substantive focus. It produced experimental materials in order to encourage further development of local adaptations and to permit revision of early materials as experience and critical review might indicate. Finally, it emphasized research and evaluation at both the national and the local level as a means of identifying sound principles and also implementing those principles in operational programs. Because of these characteristics the developmental model was both complicated and costly. How did it work?

Purely from the standpoint of logistics the CAEL developmental model worked extremely well. Active participation in the field research and trial of materials was excellent. Many individuals and institutions made very substantial contributions of time and ideas. More than most cooperative projects, the whole process of evaluating and revising assessment materials was a joint effort of many minds.

Concerning the substance of the work, it is useful to distinguish the tryout of materials from the field research phase of the work. There was less pointed criticism of experimental materials than might have been expected by such a diverse group of institutions and individuals. As a consequence relatively few major issues were raised concerning most of the original working papers. Though there were a number of improvements from the original working papers to the revised assessment reports, there were almost no fundamental redirections or serious questioning of basic assumptions underlying the assessment procedures suggested. Also, there was very limited response to CAEL's well advertised invitation to institutions to undertake the development of adaptations or modifications of specific aspects of the assessment process and have those adaptations published along with the revised reports.

Perhaps the easy interpretation is that the draft assessment materials were generally adequate and did not present any fundamental problems. This is possibly not an unreasonable assumption considering that the materials represented some of the best available examples of current practice, supplemented and strengthened by well established assessment principles from other areas of education and psychology. CAEL may have received few takers on its invitation to develop adaptations simply because most individual efforts went toward the development of local programs or toward professional publications under possibly more attractive auspices of individual scholarship. On the whole, participation in CAEL activities, extensive as it was, suggests that staff and faculty at most institutions are not likely to take strong initiative to become engaged in a large activity of this sort unless there is some direct involvement of their institution or some very specific mechanism that encourages individual participation as a visible form of professional activity.

With respect to field research it should first be recognized that there were considerable benefits of participation to the institutions that tended to further goals of the project completely aside from the primary matter of obtaining research results. A number of institutions volunteered that participation in the field research led directly to improvement in local practices, stimulated institutional research, and focused needed faculty attention on important policy and procedural issues in their institutions.

Concerning research results it was certainly expected that some findings would appear favorable and some not so favorable. One purpose of the research was to help in identifying strengths and weaknesses of assessment practices and ambiguous issues that need more attention than they have received. In the preceding pages we have seen that those were reasonable assumptions. Many of the results raise important questions that are generally applicable to educational assessment and should be matters of concern to all institutions. As such, many of these results suggest general principles pertaining to quality assurance.

At the local level there is limited evidence of systematic analyses or special studies concerning the clarity of standards, the consistency of assessment, or reasonable adherence to credit policies. Chapter VII of this report contains three exemplary studies that were carried out along the general lines of a model suggested by CAEL. These are useful examples but such studies are not yet part of the routine of most institutions. They should be because there does not appear to be any standard procedures for monitoring assessment of experiential learning that could be considered a counterpart to traditional mechanisms for monitoring the assessment and crediting of classroom learning. Because of the highly individual character of most experiential learning, traditional procedures of monitoring grading patterns, common examinations, etc. are not likely to prove sufficient.

The foregoing discussion suggests two needs. One is the need for institutions to develop the capability for routine monitoring of the assessment of experiential learning for quality assurance. That is partly a problem of integrating such assessment with traditional academic procedures for quality assurance and partly a matter of specifying periodic checks that the faculty considers important in order to insure that practices do not drift away from institutional policies and that interests of students are protected through equitable and consistent practices. Such procedures should be part of a larger conception of program evaluation that takes into account all aspects of educational effectiveness as well as costs incurred.

The second need is directly related to the first. Through its forums, discussions, committee work, and field research, CAEL has identified a number of salient practices and potential pitfalls in assessing experiential learning. Such principles of sound assessment need to be brought together for the benefit of individual institutions who could use such a statement in drafting institutional guidelines, the need for which has been cited more than once above.

It should be emphasized that any statement of principles or institutional guidelines for assessing experiential learning should be consistent with academic practices and policies generally and reflect appropriately the reality of academic assessment as it is practiced in and out of the classroom. A shortcoming of the work reported here is the fact that it may appear to treat the assessment of experiential learning as an inherently different problem than the assessment of classroom learning whereas the same principles and problems apply.

Another obvious shortcoming of CAEL's work to date is the limitation it necessarily imposed upon the scope of its work. CAEL focused upon certain problems at the acknowledged exclusion of other important issues. It has dealt with limited assessment methods, limited arenas of learning, and limited types of learning outcomes. In time those boundaries must be extended. Also CAEL has focused upon the assessment problems of higher institutions and has not yet taken into account important issues concerning educational currency; i.e., how credit for learning is recognized and transferred from institution to institution and from secondary to undergraduate to graduate and professional education.

These considerations and the foregoing discussion of findings carry implications for the style and the substance of CAEL's work in the future. While we are not here concerned with the entire field of experiential learning or even with all of CAEL's current activities, we can conclude by observing that CAEL has given much attention during the past three years to defining the substance of what is assessed, to developing assessment techniques, to establishing basic principles of good assessment, and to organizing the assessment process in an effective and responsible manner. The conclusions and implications outlined in this chapter have raised important new issues worth special emphasis.

Special attention is needed on the qualitative aspects of the assessment process, the context in which it operates, and its outcomes. We have emphasized certain problems of particular importance. These include quality assurance and program evaluation, alternate assessment methods and their costs, training needs of different types of individuals, articulation across institutions and educational levels, and how to enhance the educative character of assessment. Work on some of these problems will be especially advanced by the cooperative efforts that a continuing CAEL Assembly will make possible. One valuable activity, for example, would be cooperative studies to replicate and examine in further detail some of the questions studied here, perhaps using some of the same general research models. Such work could strengthen the interpretation of assessment principles and also assist institutions in quality assurance of assessment programs.

In addition, some of these problems will require the long-term continuing effort of individual researchers and innovative thinkers in many individual institutions. Up to now CAEL has emphasized central management of a limited number of large coordinated efforts. While these efforts have laid a necessary groundwork, an important challenge in CAEL's future is to stimulate and facilitate the continuing attention of institutions and individual professionals to the important problems and developmental needs that CAEL has helped to identify.

REFERENCES

Chen, C. C. Test analysis: College Entrance Examination Board Advanced Placement Examinations: History of art and studio art, Forms VBP, WBP, and XBP (Report Nos. SR74-33, SR75-31, and SR76-37). Princeton, New Jersey: Educational Testing Service, 1974, 1975, and 1976.

Coffman, William E. On the reliability of ratings of essay examinations. NCME Measurement in Education, 1972, vol. 3, no. 3.

Commission on Non-Traditional Study. Diversity by design. San Francisco, California: Jossey-Bass, 1973, 178 pp.

Godshalk, Fred I., Swineford, Frances, and Coffman, William E. The measurement of writing ability (Research Monograph No. 6). New York: College Entrance Examination Board, 1966, 84 pp.

Keeton, Morris T, and Associates. Experiential learning: Rationale, characteristics, and assessment. San Francisco, California: Jossey-Bass, 1976, 265 pp.

McReynolds, Paul. Reliability of ratings of research papers. American Psychologist, 1971, vol. 26, pp. 400-401.

Scott, William A. Interreferee agreement on some characteristics of manuscripts submitted to the JOURNAL OF PERSONALITY AND SOCIAL PSYCHOLOGY. American Psychologist, September 1974, vol. 29, no. 9, pp. 698-702.

Smith, Ruth. Grading the Advanced Placement English Examination. New York: College Entrance Examination Board, 1975, 38 pp.

Tinsley, Howard E. A., and Weiss, David J. Interrater reliability and agreement of subjective judgments. Journal of Counseling Psychology, 1975, vol. 22, no. 4.

Warren, Jonathan R. College grading practices: An overview. Princeton, New Jersey: Educational Testing Service, 1971, 103 pp. (Research Bulletin 71-12).

Wiggins, Jerry S. Personality and prediction: Principles of personality assessment. Reading, Massachusetts: Addison-Wesley, 1973, 656 pp.

Willingham, Warren W. CAEL project status report: March 1976. Princeton, New Jersey: CAEL, Educational Testing Service, 1976, 22 pp.

APPENDIX 1

Participants

1a. Implementation Committee

1b. Resource Panel

1c. Field Research Participants

1d. Institutions Involved in Tryout of Materials

CAEL IMPLEMENTATION COMMITTEE

Richard L. Burns
Educational Testing Service

Ruth D. Churchill
Antioch College

Arnold Fletcher
Thomas A. Edison College

Aubrey Forrest, Chairperson
Metropolitan State University

Myrna R. Miller
Vermont State College

Joseph R. Palladino
Framingham State College

Dabney Park, Jr.
Florida International University

Jose R. Rivera
El Paso Community College

Bernard Sloan
New College, University of Alabama

Urban G. Whitaker, Vice Chairperson
San Francisco State University

Warren W. Willingham
Educational Testing Service

CAEL RESOURCE PANEL

Bernice Biggs
Coordinator, General Studies
San Francisco State University

Paul Breen
Assistant Project Director
San Francisco State University

Harriet Cabell
Associate Director
External Degree Program
New College, University of Alabama

Anne Calhoun
Assistant Professor
Metropolitan State University

Ruth Churchill
Dean for Educational Evaluation
Antioch College

Laurent Daloz
Director of Learning Services
Community College of Vermont

John S. Duley
Director, Field Experience Program
Justin Morrill College
Michigan State University

Arnold Fletcher
Vice President for Academic Affairs
Thomas A. Edison College

Aubrey Forrest
Director of Assessment
Metropolitan State University

Sheila Gordon
Associate Dean for Cooperative Education
LaGuardia Community College

Jan Hagberg
Assistant Professor
Metropolitan State University

Leah Harvey
Instructor, Office of Assessment
Metropolitan State University

Eugene J. Kray
Dean of Community Education
Delaware County Community College

Peter Meyer
Professor of Social Work
Florida International University

Myrna Miller
Coordinator, Assessment Services
Community College of Vermont

Dabney Park, Jr.
Director, External Degree Program
Florida International University

Judith Pendergrass
Director of Admissions and Records
Metropolitan State University

Martha Sachs
Director of Adult Education
Ramapo College of New Jersey

Robert F. Sexton
Executive Director
Office for Experiential Education
University of Kentucky

Mahesh Sharma
Acting Coordinator for Evaluation
Educational Development Center

Bernard Sloan
Coordinator, Out-of-Class Learning
New College, University of Alabama

Urban G. Whitaker
Dean of Undergraduate Studies
San Francisco State University

The following ETS staff collaborated on CAEL developmental work during 1974-75:

Thomas F. Donlon
Joan E. Knapp
Hadley S. Nesbitt
Jane Porter
Richard R. Reilly
Amiel T. Sharon

1-3

INSTITUTIONS PARTICIPATING IN CAEL FIELD RESEARCH

	AREA			
INSTITUTION	A	B	C	D
Alverno College	X			X
Antioch College				X
Community College of Vermont		X		
Delaware County Community College			X	
Empire State College		X		
Governors State University		X		X
Hartwick College	X			X
LaGuardia Community College			X	
Macalester College	X			
Metropolitan College, St. Louis University		X		X
Metropolitan State University	X			
New College, University of Alabama		X		X
North Carolina Agricultural and Technical State University			X	
Notre Dame College			X	
San Francisco State University	X	X		
Staten Island Community College	X	X	X	
Stephens College	X			
Texas Christian University		X		
Thomas A. Edison College				X
University of Akron	X		X	X
University of California at L. A.	X			
University of Missouri		X		X
University of Oregon	X	X		X
Wichita State University			X	

APPENDIX 1d

INSTITUTIONS INVOLVED IN TRYOUT OF MATERIALS

Alverno College
American College
Antioch College
Augustana College
Berea College
Bethany College
Black Hawk College
Brevard College
California State College at Bakersfield
California State University at Los Angeles
Central Michigan University
College of Saint Benedict
College of Saint Rose
Community College of Vermont
Delaware County Community College
Delta College
Eckerd College
El Paso Community College
Empire State College
Everett Community College
Florida International University
Fort Wright College
Golden West College
Governors State University
Hartwick College
Johns Hopkins University
Johnston College, University of Redlands
Justin Morrill College, Michigan State University
Kansas State University
Kent State University
Kentucky State University
King's College
LaGuardia Community College
Lone Mountain College
Luzerne County Community College
Macalester College
Madonna College
Metropolitan College, Saint Louis University
Metropolitan State College
Metropolitan State University
Miami-Dade Community College
Mount Mary College
Mount Union College
New College, University of Alabama
North Carolina Agricultural and Technical State University
Notre Dame College
Our Lady of Angels College
Our Lady of the Lake College
Rochester Institute of Technology
Roosevelt University
Sacred Heart University
Saint Thomas Aquinas College
San Francisco State University
Southern Illinois University
Southern Oregon State College
State University College at Brockport
Staten Island Community College
Stephens College
Texas Christian University
Texas Southern University
Thomas A. Edison College
Towson State College
Union College
University of Akron
University of California at Los Angeles
University of Cincinnati
University of Dayton
University of Evansville
University of Kentucky
University of Maryland
University of Massachusetts
University of Minnesota
University of Missouri
University of Oregon
Virginia Polytechnic Institute and State University
Whatcom Community College
Wichita State University
Winona State University

APPENDIX 2

Background Information

2a. Prospectus

2b. Invitation to Participate in CAEL Activities and Tryout of Materials

2c. Schedule for CAEL Validation Work

2d. Initial Communication with Participants

2e. General Instructions to Coordinators

2f. Agenda for January 14 Meeting With Site Coordinators

2g. Request for Feedback on Tryout of Materials

2h. Request for Reactions to Tentative Results

2i. Working Papers and Institutional Reports (1975)

2j. Revised Reports (1976)

cael

PROSPECTUS

for Institutional Participation in the tryout of CAEL Materials

COOPERATIVE ASSESSMENT
OF EXPERIENTIAL LEARNING

This Prospectus describes how institutional members of the CAEL Assembly can participate during 1975-76 in the experimental tryout of CAEL materials for assessment of experiential learning. Institutional participation will be coordinated through each institution's representative to the Assembly. In early June representatives will receive further information and forms for (1) indicating what specific options for participation the institution prefers and (2) applying for special CAEL programs that may permit only limited participation.

April 1975

Individuals seeking answers to specific questions or wishing to have their names added to the CAEL mailing list for periodic information about the project should write to CAEL, Educational Testing Service, Princeton, NJ 08540.

Additional copies are available, without charge, upon request.

Copyright © 1975 by Educational Testing Service. All rights reserved.

CAEL PROSPECTUS
April 1975

Warren W. Willingham

BACKGROUND

The CAEL Project is a cooperative undertaking that involves Educational Testing Service, ten Task Force Institutions, and a much larger CAEL Assembly of member institutions. The purpose of the project is to develop and try out materials and procedures that will be useful in strengthening the assessment of experiential learning in higher education. (See *Status Report: December 1974* for a more detailed description of CAEL.) In its first year the project focused developmental work on seven tasks in four priority areas. The results of that work were reported in the following seven Working Papers in the spring of 1975:

AREA A — Assessment of Interpersonal Skills

Working Paper 4. The Learning and Assessment of Interpersonal Skills: Guidelines for Administrators and Faculty.
Working Paper 5. The Learning and Assessment of Interpersonal Skills: Guidelines for Students.

AREA B — Use of the Portfolio to Assess Non-Sponsored Learning

Working Paper 6. A Guide for Assessing Prior Experience through Portfolios.
Working Paper 7. A Student Handbook on Preparing a Portfolio for the Assessment of Prior Experiential Learning.

AREA C — Assessment of Learning Outcomes of Work Experience

Working Paper 8. A Task-Based Model for Assessing Work Experience.
Working Paper 9. A Student Guide to Learning through College-Sponsored Work Experience.

AREA D — Use of Expert Judgment to Assess Learning Outcomes

Working Paper 10. The Use of Expert Judgment in the Assessment of Experiential Learning.

These working papers include descriptions of competences, suggested assessment procedures, specimen materials, and so on. During 1975-76 the developmental work moves into a second important phase—the tryout of these experimental materials in CAEL Assembly institutions. The Assembly includes a broad cross-section of American higher education, and the plan for directly involving these institutions in the validation of developmental work represents a unique model for insuring that CAEL outcomes are sound and for facilitating direct benefit of CAEL work in many institutions.

The participation of as many Assembly institutions as possible in the tryout of these experimental materials during 1975-76 is vital for several reasons. First, it is important to review critically experimental materials in a number of institutional settings. Second, various studies will be necesssary in order to demonstrate reliability and validity of the assessment methods. Third, CAEL needs systematically to encourage modification and adaptation of prototype procedures to a variety of different types of students and learning situations, thereby broadening applicability and usefulness without standardizing assessment procedures. This is a very important feature of the work because individualized assessment cannot be routinized. At the same time, a major goal of the CAEL project is to affect beneficially assessment practices in a much larger circle of institutions than those heavily committed to nontraditional learning experiences and programs. The purpose of this *prospectus* is: I. To give an overview of the different types of assessment materials that are available for experimental use. II. To describe briefly how the tryout plan will work—the advantages and conditions of participation, the general options available, etc.

I BRIEF DESCRIPTION OF CAEL ASSESSMENT MATERIALS IN FOUR PRIORITY AREAS

AREA A—Assessment of Interpersonal Skills

Work in Area A has centered on the development of two manuals—one for faculty and one for students. The primary focus of the manuals reflects an analysis of the interpersonal domain originally due to Sidney Fine. Dr. Fine and his associates developed a classification system for occupational roles, differentiating them according to their demands for work with people, with data, or with things. Within this classification he suggested a number of subcategories, and those relating to people have been taken as a basis for a CAEL analysis. The greatest value to CAEL in the adaptation and extension of this existing system lies perhaps in the focus on the important societal roles which demand interpersonal ability.

The faculty guide provides a general framework for understanding interpersonal skills and their relation to undergraduate programs. The student manual translates these concepts to the student's point of view and provides illustrative case studies from institutions offering experientially based credit for interpersonal competencies.

In addition to a theoretical structure these two working papers provide a description of strategies for assessment based on prior work presented in CAEL Working Paper No. 2. The systematic survey of assessment approaches which was presented there has been retained and the applications of these methods to problems in the interpersonal domain are explored. Furthermore a general purpose model is suggested that includes alternate procedures for adaptation to particular assessment situations on individual campuses.

In addition to review and critique, possibilities for institutional validation and use of these materials during 1975-76 include:

- experimental use of the student guide
- adaptation of the model assessment procedure to local needs
- development of one of the suggested assessment techniques (e.g., leaderless group discussion) to a particular experiential learning program
- evaluation of the validity of interview assessments in relation to field observation of interpersonal skill.

AREA B—Use of the Portfolio to Assess Non-Sponsored Learning

This area includes two Working Papers: a faculty guidebook and a student guide. The faculty guidebook attempts to answer three questions that faculty might raise before embarking on the assessment of prior (or non-sponsored) learning via the portfolio: 1) What are the types of prior experiences and learning outcomes resulting from these experiences that will be presented to faculty for assessment? 2) What portfolio processes are now used to assess prior learning? 3) What choices do faculty have concerning procedures that might be used in portfolio assessment?

The guide addresses the first two questions by providing information about the state-of-the-art in portfolio assessment. Approximately 400 portfolios were examined during more than 30 institutional site visits. The third question is addressed by describing alternate procedures in each stage of a model procedure of portfolio assessment. Sample checklist forms, outlines, instruments from cooperating institutions are included to illustrate various procedures. Advantages and disadvantages of alternate procedures are also discussed.

The student workbook is written in an informal style and is designed for use by adult students entering or returning to higher education. It is written in such a manner so as to take into account a variety of institutional assessment procedures and arrangements. The situations range from institutions that have no special program for prior learning and assessment to institutions that have credit courses in educational planning and learning assessment. It describes clearly and simply the steps in portfolio assessment and uses illustrations of six students who are seeking university degrees. Prior or life experience is described for each of the students, and the workbook user is set to the task of applying the portfolio assessment procedures to the examples and eventually to him/her. This document is written not only to help the student with these practical problems concerning the construction of a portfolio but is designed to guide the student through what could be a highly educational experience.

In addition to review and critique, possibilities for institutional validation and use of these materials during 1975-76 include:

- experimental use of the student guide
- adaptation of the model portfolio procedure by systematically considering and selecting alternate procedures and developing necessary supporting documents
- determine the extent of faculty agreement as to what constitutes creditable learning in various areas of non-sponsored learning
- study the accuracy of various types of documentation.

AREA C—Assessing the Learning Outcomes of Work Experience

This area includes two pieces of work, each of which can stand alone. The two lines of work are described in two reports: Working Paper No. 8., A Task-Based Model for Assessing Work Experience, and Working Paper No. 9, A Student Guide to Learning through College-Sponsored Work Experience.

Working Paper No. 8 presents a generalizable model for assessing specific competencies in occupational fields, especially competencies required in career related degree programs. The model is designed to help a college specify the kinds of skills and competencies acquired in various occupational settings, to define the learning outcomes of its occupational and career programs, and to translate students' relevant work experience into college credit.

A competency is defined in this model as the ability to perform a very specific task in a specific occupational context, e.g., operate radar to identify violators of speed laws, prepare detail program flow charts or decision tables, and interpret assembly drawings and layout details when no detail drawings are available. In seven sequential stages the model describes how lists of such competencies can be developed in an occupational field and how they can be used for various objectives in higher education which require the assessment of work experience.

Working Paper No. 8 includes a detailed application of the Work Assessment Model to the fields of data processing, law enforcement, and secretarial science. In each of these fields a number of prototype assessment instruments or procedures have been developed on the basis of compendia of competencies. Thus, in these three fields there are products that could be immediately applied in a practical situation.

It is anticipated that the model will be most useful to institutions that offer occupational and career programs. Since the assessment and crediting of work experience in occupations is most often performed at the departmental level, the working paper is likely to be helpful to department chairmen and other subject matter experts who are actually involved in the assessment of work experience within the college.

Working Paper No. 9 is a guide for students who are participating or are planning to participate in sponsored work experience programs (work-study programs, internships, cooperative programs, etc.). It can serve as a combined planning guide, study outline, and self-evaluation device, to help students meaningfully integrate their work experience into their overall educational programs.

The student guide discusses such issues as career exploration, acceptance of responsibility, preparation for the learning experience, involvement in the experience, and synthesis of the learning following the experience. It recognizes differences between various types of off-campus programs, discusses the use of self-ratings and other ratings, and directs students to appropriate information sources on their own campuses.

More specifically, the guide outlines a series of eleven steps through which the student is encouraged to progress. These range from taking stock of goals and considering available work options to assessing learning outcomes after the off-campus experience and identifying new learning objectives. These steps are accompanied by appropriate worksheets and checklists. Although the guide is designed so that it can be used in its entirety in a variety of settings, each of its sections will be adaptable to individual usage in cases where faculty might prefer to substitute their own materials for portions of the guide.

In addition to review and critique, possibilities for institutional validation and use of these materials during 1975-76 include:

- experimental use of the student guide
- adaptation of the work assessment model to a local program in one of the three areas illustrated in the guide
- development of a similar work assessment model in a different field
- examine the extent to which learning credited through such work assessment is comparable to parallel learning within the degree program.

AREA D—The Use of Expert Judgment in Assessing Learning Outcomes

Due to the highly individual character of much experiential learning, assessment often must rely upon the informed judgment of an expert. Working Paper No. 10 provides principles and guidelines for faculty, students and consultants on the use of expert judgment. The first portion of the report deals with procedural issues that cut across the various contexts in which expert judgment might be applied. Several topics receive emphasis.

First, the role of the expert in assessment includes a number of separate functions in addition to actually judging competence. Such functions as defining criteria and structuring and adapting the assessment procedure to the student are equal in importance to the function of judgment and in fact make reliable, valid judgment possible. Second, the importance of establishing standards for assessment is emphasized. Standards should be defined as systematically as possible and related to objective, observable indices or behaviors. Third, the issues of reliability and validity are illustrated with examples of rating errors that bear on the credibility and the fairness of the assessment. Evaluation of assessment is also treated; i.e., suggestions as to how assessment might itself be assessed. Finally, general principles and steps in sound assessment are outlined.

These various issues, principles, and steps in assessment are illustrated and discussed in a series of chapters covering the use of judgment in four areas: interviews, product assessment, performance assessment, and assessment of written material. Discussion of each of these four areas in which expert judgment is commonly applied will include a number of practical hints and suggestions for the improvement of assessment as well as problems and pitfalls to avoid. Several applications are described in some detail; e.g., prototype assessment interviews, a model for product assessment in the visual arts, an institutional plan for performance assessment, and an assessment model for off-campus experiential learning programs.

In addition to review and critique, possibilities for institutional validation and use of these materials in 1975-76 include:

- local adaptation and application of any of the various types of assessment procedures outlined in the report; i.e., the structured interview, the performance assessment plan, the visual arts assessment model, or the procedure for assessing off-campus learning
- studies of the reliability of interview assessment (agreement among experts)
- exemplary projects to develop criterion standards that can be applied locally to student, institutional, or employer goals
- validity studies that establish the comparability of expert assessment to what students are actually able to do in real situations.

II THE TRYOUT PLAN: THREE LEVELS OF PARTICIPATION

Since the CAEL Assembly includes a diversity of colleges, universities, and agencies, it seems certain that different Assembly members will profit from CAEL experimental materials in different ways. Assembly members that do not yet have active programs of experiential learning may wish to use CAEL materials as a basis for staff discussion and planning. In the preparation of assessment materials the project staff and consultants have identified many issues of importance to institutions that are considering the introduction or expansion of experiential learning programs. Other member institutions that already have programs in areas addressed by CAEL can use student and faculty guides experimentally (gratis for limited copies) as a basis for strengthening assessment procedures. Still other institutions may wish a more substantial involvement with the project and decide to volunteer for field research. This latter alternative offers more intensive contact with CAEL staff through workshops or special activities and can help to stimulate more interest on the campus through participation in research and development or self-study models.

These three levels of participation in the validation work—1) Review and Critique, 2) Experimental Use, and 3) Field Research—are supplementary in the sense that involvement at Level 2 automatically includes review and critique, and involvement at Level 3 necessarily involves experimental use of materials. The three levels of participation are described below with respect to their purposes and what is likely to be involved.

The general plan will work as follows. The Working Papers will form the basis for several workshops at the May 1975 meeting of the Assembly. (These workshops will be repeated for new members as necessary at the fall meeting of the Assembly.) Member institutions should select which workshops to attend according to the developmental work of greatest interest to them. The purpose of these several-hour workshops will be to discuss in some detail the plans for institutional participation in validation of particular materials during 1975-76. Within a few weeks following these workshops, each institution will be asked to indicate its interest in participating in the tryout of CAEL materials. At that time the institution will choose one or more priority areas and indicate the level of participation desired in each case. The three levels of institutional involvement can be described briefly as follows.

Level 1 Participation: Review and Critique

A principal purpose of the validation phase of CAEL is to review carefully the experimental assessment materials that are produced during 1974-75. This is necessary not only to spot and correct weaknesses, but also to identify additional issues, problems, and developmental possibilities that deserve attention. This critical review should involve students and faculty from all types of institutions as well as others who may have some responsibility in fostering experiential learning of students.

It is likely that most CAEL Assembly members will wish to undertake a systematic review of those materials that are of special interest to them. Institutions that are interested in a particular priority area but do not have an operating program (e.g., the institution has no provision for crediting prior experiential learning) would necessarily have to restrict their participation in the tryout of such materials to a critical review by faculty and students. The work that has gone into preparing the CAEL materials can be of direct benefit to institutions or departments considering new activities or programs involving experiential learning. Systematic review of CAEL materials by a local committee could be an effective way to clarify issues, focus discussion, and aid program planning.

This level of participation in the validation plan would not necessarily involve direct contact between the institution and individuals working on CAEL development except through regular CAEL Assembly meetings. It will involve, however, a request that individual institutions report in writing to CAEL by March 15, 1976 the results of their critical review and any recommendations they can make regarding needed revisions or extension of the work.

Level 2 Participation: Experimental Use

A second major purpose of the validation phase of CAEL is to try out the experimental materials in a number of quite different situations. This is partly to test the appropriateness of suggested procedures but also to encourage systematic modification and adaptation of general assessment procedures and prototype materials to the special conditions of different learning situations and populations of students. CAEL actively encourages the development of local assessment models in order to enhance the appropriateness of assessment procedures and to avoid undesirable standardization of techniques or instruments.

This level of participation will be open to institutions that have an active program in which materials can actually be tried out with students. Naturally such experimental use should also involve systematic review and critique, but more is implied. Many institutions using CAEL materials on an experimental basis will wish to consider systematically how those experimental procedures and materials might best be modified to meet local needs and circumstances while maintaining the suggested principles of sound and equitable assessment. Ideally, this sort of systematic examination and adaptation should result in a specially designed local assessment model that could be described in written form for the benefit of all CAEL institutions. CAEL hopes to incorporate some selected institutional reports of this kind in the revised assessment materials as exemplary applications. At the least, the institution's experience with the experimental materials should be reported back in writing to the project staff at the end of the tryout period (March 15, 1976).

During the experimental period contact would be maintained between CAEL staff and participating institutions through written communication as necessary. CAEL will provide limited assessment materials for experimental use at no charge to institutions involved in such tryout. In summary, this level of participation involves an institutional agreement to try out CAEL materials and to report back either (1) reactions to the experimental tryout, or if the institution wishes, (2) a more formal description of how the assessment materials were adopted for local use.

Level 3 Participation: Field Research

A third major purpose of the validation phase of CAEL is to carry out the necessary research and special studies to insure that suggested assessment procedures are technically adequate and also workable. In particular, it is important to give proper attention to validity and reliability so that assessment techniques fostered by CAEL are characterized by consistency and equity.

Institutions that are heavily involved in experiential learning programs are especially concerned about such questions and the quality of assessment in general. Many of these institutions have sufficient students in experiential learning programs to make possible special studies that require gathering information from students. These institutions are urged to volunteer as field research sites in order to participate more fully in the validation of CAEL materials. Institutions will be selected from among those that express an interest in being a research site.

In addition to experimental use of materials this level of participation will likely involve systematic record-keeping for a group of students, or questionnaire studies, or special evaluation of procedures, and so on. This sort of cooperative work will likely require special routine communication and meetings with representatives from research sites. The expected outcome will be staff reports of research that help to clarify what assessment procedures are sound from a technical standpoint and how assessment can be strengthened. Institutions can expect to profit from this arrangement through involvement in additional CAEL activities and workshops concerned with problems of developing sound local assessment procedures.

Depending upon level of CAEL funding from foundation sources for 1975-77, some of this field activity may involve programs of faculty development or special opportunities for institutions to collaborate in working on institutional problems of implementing assessment programs. Such activities will depend largely upon institutional commitment and released staff time for local work in cooperation with other CAEL institutions though very limited funds may be available to offset some direct clerical expenses incurred by institutions.

All of these types of field activity at the third level of participation in the tryout of CAEL materials will afford a limited number of selected institutions a unique opportunity to engage students and key faculty in an intensive, structured examination of assessment procedures and related aspects of experiential learning programs. More specific information about the nature of participation at this third level and the options available will be included in a mailing to institutional representatives in early June—at which time CAEL will request institutional preferences. In August an external committee will select institutions for intensive involvement from among those that express such interests.

NEXT STEPS

To summarize briefly, the CAEL project has produced during 1974-75 a series of Working Papers that incorporate assessment materials in four broad areas:

A. Assessment of Interpersonal Skills
B. Use of the Portfolio to Assess Non-Sponsored Learning
C. Assessment of the Learning Outcomes of Work Experience
D. Use of Expert Judgment to Assess Learning Outcomes

This prospectus outlines a general plan for trying out these materials in institutions of the CAEL Assembly during 1975-76. Institutions can participate at three levels:

I. Review and Critique
II. Experimental Use
III. Field Research

What next steps are appropriate in considering such possibilities, and what sequence of events can institutions anticipate?

1. After going through the materials, it would be helpful for institutional representatives to share this Prospectus, the December Status Report, and selected Working Papers with other interested persons on the campus who might like to be involved in tryout work and the decision to do so. Discussion with colleagues should be very helpful in reaching an initial, tentative decision regarding the area(s) of greatest interest.

2. Representatives should give special attention to those areas at the May Assembly meeting. It is important to attend the workshops in which tryout of the relevant materials is discussed. A special session is scheduled for institutions interested in Level III participation. (Institutions joining the CAEL Assembly after the meeting in May will be briefed at a fall meeting regarding participation in the tryout, but Level III participation will require meeting the July 15 application deadline as described below.)

3. Following the Assembly meeting institutional representatives should brief colleagues on their campus as necessary. In early June, CAEL will send a form requesting specific preferences from member institutions regarding tryout participation. One basic piece of information requested at that time will look something like this:

	Area A Interpersonal Skills	Area B Portfolio	Area C Work Experience	Area D Expert Judgment
Review & Critique	I	[✓ I]	I	I
Experimental Use	II	II	II	II
Field Research	III	III	[✓ III]	III

For any area in which the institution would like to participate, the respondent will be asked to circle I, II, or III in the appropriate column. For example, the circles marked above would indicate that the institution is interested in active participation in two areas. It wishes to review and critique the materials concerning use of the portfolio. It also wishes a much more substantial involvement in the area of assessing work experience and is willing to commit staff time to local work in cooperation with CAEL.

4. All institutions designating Level I or II can be accommodated. Participation at Level III will be limited. Member institutions will be asked to reply by July 15. Those interested in Level I or II can reply later if necessary, but those applying for participation as a field research site must do so by July 15. There may be several types of such field research activity; these will be specified in the June mailing.

5. About August 1 a special external committee will review applications for cooperative field work and select the institutions that will participate at Level III.

6. During August representatives can expect to hear from CAEL regarding what institutions will be involved in various activities, plans for special meetings, etc.

APPENDIX 2b

cael
COOPERATIVE ASSESSMENT
OF EXPERIENTIAL LEARNING

A JOINT PROJECT OF EDUCATIONAL TESTING SERVICE AND A GROUP OF COLLEGES AND UNIVERSITIES WITH INITIAL FUNDING BY THE CARNEGIE CORPORATION.

ETS, Princeton, New Jersey 08540 ▪ 609-921-9000

NOTE: Enclosures concerning operational models and faculty development are not included in this Appendix.

Morris T. Keeton, *Chairperson*
Provost and Vice President
Antioch College

Neal R. Berte
Dean, New College
University of Alabama

James D. Brown Jr.
President
Thomas A. Edison College

Arthur W. Chickering
Vice President for Policy
Analysis and Evaluation
Empire State College

Lawrence E. Dennis
Provost and Director
Massachusetts State College
System

Cyril O. Houle
Professor of Education
University of Chicago

Winton H. Manning
Vice President, Higher Education
and Career Programs
Educational Testing Service

Jacquelyn A. Mattfeld
Dean of Academic Affairs
Brown University

Jules O. Pagano,
Vice Chairperson
Associate Vice President
Office of Community Affairs
Florida International University

Jean M. Pennington
Director, Continuing Education
for Women
Washington University

David H. Provost
Dean, Academic Planning
and Evaluation
California State University
and Colleges

Gilberto de los Santos
Dean of Instruction for
Research and Development
El Paso Community College

Herman B. Smith
Director, Office for the
Advancement of Public
Negro Colleges
National Association of
State Universities
and Land Grant Colleges

Peter B. Smith, *Secretary*
Director
Community College of Vermont

David E. Sweet
President
Minnesota Metropolitan State
College

▪

Warren W. Willingham
Executive Director for Program
Research
Educational Testing Service,
Project Director

TO: REPRESENTATIVES OF CAEL MEMBER INSTITUTIONS FOR 1975-76

FROM: Warren W. Willingham

DATE: June 6, 1975

SUBJECT: Invitation to participate in CAEL activities and tryout of materials

There are several types of activities in which CAEL member institutions can participate during 1975-76. The enclosed Prospectus differentiates three levels of participation:

Level 1 - Review and Critique (all members eligible)

Level 2 - Experimental Use (all members eligible)

Level 3 - a) Field Research (limited to 30 institutions)
- b) Faculty Development Program (limited to 12 institutions)
- c) Development of Operational Models (limited to 12 institutions)

Level 1 and 2 are open to all member institutions. The green packet enclosed summarizes briefly these levels of participation and provides an application blank for those that want to be so engaged in CAEL work next year. Institutions interested in using CAEL materials experimentally should return the green form regardless of whether you are also applying for Level 3 involvement.

Interested institutions are urged to apply for the more substantive involvement of Level 3—especially the field research that is so necessary to validate CAEL assessment materials. The three types of Level 3 involvement are described in the three white packets of information. Institutions interested in applying to more than one of these three must send in each appropriate application sheet (e.g., an application to be a CAEL Field Research Site as well as an application to participate in the faculty development program if both are desired).

Institutions can apply for Level 1 or 2 (green form) at any time through fall 1975, but white applications for Level 3 participation must be mailed by July 15 because a special committee will select those institutions in mid-summer. Institutions selected will be notified in August and further information regarding these activities will be available by that time and will be forwarded to those individuals involved.

Please direct applications and inquiries regarding these activities to CAEL, Educational Testing Service, Princeton, New Jersey 08540, attention Ms. Jane Porter, telephone (609) 921-9000, extension 2186.

Institutional Participation in the Tryout of CAEL Materials: Review and Critique and Experimental Use

Level I: Review and Critique[*]

During the coming year the CAEL project must carry forward and complete a great deal of developmental work that has proceeded to the point where it can be shared with confidence and used experimentally. While the Steering Committee regards this initial work as having clear value for operational use, it recognizes that the procedures and materials now being readied for tryout can be improved in many ways, and that part of that improvement is the additional description of model applications at various institutions of the more general assessment methods described. So an important outcome of the validation phase is the revision, improvement and actual production of formal published materials.

An obvious problem lies in the fact that CAEL is not producing a set of standard instruments that are routinely subjected to statistical criteria of technical adequacy. As has been its intention CAEL is producing detailed principles, guidelines, suggested procedures, prototype documents, and methods that are generally applicable to sound assessment in an indefinite variety of learning situations. There is no way that a few technical studies can establish the validity and reliability of CAEL materials in the sense that computation of certain correlation coefficients shows a standardized test to be reliable and valid for its purpose. There are, however, important ways in which the appropriateness and validity of CAEL materials and procedures need to be examined, demonstrated, and supported.

One way in which this can be accomplished is through a systematic process of critical review by all parties likely to be concerned with or affected by assessment.

A principal purpose of the validation phase of CAEL is to review carefully the experimental assessment materials that are produced during 1974-75. This is necessary not only to spot and correct weaknesses, but also to identify additional issues, problems, and developmental possibilities that deserve attention. This critical review should involve students and faculty from all types of institutions as well as others who may have some responsibility in fostering experiential learning of students.

It is likely that most CAEL Assembly members will wish to undertake a systematic review of those materials that are of special interest. Institutions that are interested in a particular priority area but do not have an ongoing program (e.g., crediting prior learning) would necessarily have to restrict their participation in the tryout of such materials to a critical review by faculty and students. For institutions or departments considering new activities or programs involving experiential learning, systematic review of CAEL materials by a local committee could be an effective way to focus discussion and aid program planning.

---

[*] This brief overview is abstracted from the Prospectus.

This level of participation in the validation plan would not necessarily involve any extensive contact between the institution and individuals working on CAEL development except through regular CAEL Assembly meetings. It would involve, however, a request that individual institutions report to CAEL the results of their critical review and any recommendations they can make regarding revisions or extension of the work.

Schedule and Conditions of Participation for Level I

The schedule calls for feedback from institutions by March 1976. The three following formal conditions have been established for participation at Level I.

A. CAEL Assembly members have the option of undertaking a formal review and critique of CAEL assessment materials in one or more of the four developmental areas (A - Interpersonal Skills; B - Use of Portfolio to Assess Non-Sponsored Learning; C - Work Experience; D - Expert Judgment).

B. In any area so selected for review and critique CAEL will provide the institution with gratis copies of the relevant Working Papers as may be necessary (up to a limit of five faculty guides and ten student guides).

C. The institution will request interested faculty and students to respond to some specific questions about the adequacy of the materials they review (i.e., return a brief CAEL questionnaire in spring 1976).

Level II Participation: Experimental Use

A second major purpose of the validation phase of CAEL is to try out the experimental materials in a number of quite different situations. This is partly to test the appropriateness of suggested procedures but also to encourage systematic modification and adaptation of general assessment procedures and prototype materials to the special conditions of different learning situations and populations of students. CAEL actively encourages the development of local models in order to enhance the appropriateness of assessment procedures and to avoid undesirable standardization of techniques or instruments.

This level of participation is open to institutions that have an active program in which materials can actually be tried out. Naturally such experimental use should involve systematic review and critique, but more is implied. Institutions using CAEL materials on an experimental basis will wish to consider systematically how those experimental procedures and materials might best be modified to meet local needs and circumstances while maintaining principles of sound and equitable assessment. Ideally, this sort of systematic examination and adaptation should result in a specially designed local model that could be described in written form for the benefit of all CAEL institutions. In any event the institution's experience with the experimental materials should be reported back to the project staff at the end of the tryout period.

Schedule and Conditions of Participation for Level II

Institutions are expected to report back formally the results of their adaption by March 1976. The conditions of participation are:

A. CAEL Assembly members have the option of using CAEL materials experimentally during 1975-76 in any developmental area in which the institution has an active program involving students.

B. In any area so selected for experimental tryout of materials, CAEL will provide the institution with gratis copies of the relevant Working Papers as may be necessary (up to a limit of 10 faculty guides and 25 student guides).

C. CAEL advocates institutional adaptation of assessment procedures to the conditions and needs of local programs. Any institution developing such an adaptation is encouraged to write it up as an institutional model and submit it to CAEL for possible use as an illustration of an exemplary procedure. Any such reports should be sent to CAEL by March 15, 1976. At the least, institutions will be requested to provide written feedback similar to that specified under Level I.

How to Apply. Institutions wishing to participate at Levels I and/or II should return the attached application indicating levels and the areas of interest. Note that institutions are not limited to one area. Institutions are asked to return the attached sheet by July 15, 1975, if possible, though later receipt of this particular application is acceptable.

Application for Participation in the Tryout of CAEL Materials

(Levels 1 and 2)

This institution has applied for 1975-76 membership in the CAEL Assembly. We would like to participate in the tryout of CAEL materials in the areas indicated below. (Place an X under Level 1 or Level 2 for each area in which you wish to participate. Also please indicate the number of faculty and student participants you expect so that materials can be forwarded.)

		Level of Participation		No. of Participants	
		1. Review & Critique	2. Experimental Use	Faculty	Student
Area A.	Assessment of Interpersonal Skill	______	______	______	______
Area B.	Use of Portfolio to Assess Non-Sponsored Learning	______	______	______	______
Area C.	Assessment of Work Experience	______	______	______	______
Area D.	Use of Expert Judgment in Assessment	______	______	______	

Institutional Representative ________________________________

Name of Institution ________________________________

Telephone ________________________________

Return to: CAEL
Educational Testing Service
Princeton, New Jersey 08540
Attention: Jane Porter

Level 3 Participation:
Involvement in Field Research

During 1975-76 CAEL will need the cooperation of some 25-30 member institutions as Field Research Sites for the validation of assessment materials. These institutions will participate in studies that will require some faculty and student time but should be intrinsically interesting and useful to those concerned with assessment of experiential learning. The following section describes the nature of this participation ($2000 per institution is provided to cover local clerical and miscellaneous expenses that may be incurred). All interested members are urged to participate; to do so, send in the application sheet attached to this packet.

Description of Field Research

During the coming year CAEL materials must be validated in two major ways. The first way is through a process of systematic review, critique and local adaptation of the materials. This general type of involvement (described as Level 1 or 2 in a separate enclosure) will provide important feedback on the usefulness and appropriateness of CAEL materials.

The second major way in which institutions are urged to participate in validation is through involvement in field research. This aspect of the validation work refers to the process of verifying that the assessment materials and procedures are adequate from an educational and assessment standpoint; that is, that they have satisfactory validity, reliability, and educational benefit.

This can be accomplished; (a) through studies that reveal important characteristics of assessment procedures that tend to enhance the reliability and validity of local applications; and (b) through the development of validation models whereby individual institutions can monitor the quality control of assessment procedures, especially with respect to the consistency of assessment and to achieving consensus regarding its content and process.

The studies in Areas A, B, and C below each focus upon particular educational and psychometric problems at different stages of assessment (i.e., one or more of the six stages frequently referred to in CAEL reports). Each of these studies (A1 through C4) will involve collection of data from several institutions in accordance with an overall study design along the lines indicated below. The CAEL staff will analyze the data and report results in one or more CAEL technical reports, fully protecting anonymity of individual and institutions. Institutions that wish to participate as a CAEL field research site will agree to participate in all of the studies in one of the areas - A, B, or C (or more than one area if the institution wishes).

Area D includes a different type of study that is optional for all CAEL field sites. Thus, if a field site institution wishes it can undertake a demonstration validation of expert judgment (Study D1) in addition to participating in Area A, B, or C. These demonstration validity studies will be carried out and reported by the institution within the general framework of a procedure and set of suggestions provided by CAEL.

AREA A - INTERPERSONAL SKILLS

Stage 2:
Construct
Validity

Study #A1 - To what extent can faculty articulate interpersonal competencies with different degree programs?

Rationale: If there is a clear rationale for crediting interpersonal skills, then there should be acceptable agreement among the faculty on the differences among degree programs with respect to which interpersonal competencies are relevant.

Data: The basic data will be judgments of faculty at field research sites as to which interpersonal competencies are relevant to each of several selected degree programs. The interpersonal competencies to be judged will be presented in the form of approximately 20 illustrative cases.

Requirements: Each research site must provide a complete set of faculty judgments for the illustrative cases. A minimum of five faculty per research site will be required.

Stage 4:
Reliability

Study #A2 - To what extent do different assessors agree on the level of interpersonal competence students have acquired?

Rationale: In order for assessment of interpersonal skills to be fair and valid there should be reasonable agreement between different assessors using the same method and between assessments made using different methods.

Data: At least two independent assessments of the level of interpersonal competence will be made for a sample of students using a structured interview and a sample of behavior (recorded or live).

Requirements: Field research sites must assess a minimum of 20 students with an interview and a taped performance sample, in each case providing two independent assessments.

All Stages:
Educative
Value of
Assessment

Study #A3 - What is the educational value to the student of the assessment of interpersonal skills?

Rationale: If assessment of interpersonal skills is of educative value then a significant improvement should be observed in students' personal growth and development before and after assessment.

Data: Semi-structured free response measures relating to student understanding of their competencies and their educational goals will be obtained for a sample of students prior to assessment and for another sample of students after assessment has taken place.

Requirements: Institutions must be willing to provide such information from a minimum of 20 students randomly divided into those for whom information is collected prior to assessment and those for whom information is collected after assessment. Institutions must also be willing to employ an approximate version of the model for assessing interpersonal skills as described in CAEL Working Paper No. 4.

AREA B - PORTFOLIO ASSESSMENT

Stage 5:
Reliability
of Standards

Study #B1 - Do faculty agree on the amount of credit awarded for portfolios?

Rationale: In order to be fair and valid, crediting practices must be consistent. One way of assessing the consistency of crediting practices is to see how closely assessors agree on the number of credits to be awarded after independently examining the same evidence.

Data: Two different types of data will be collected. The first type will be actual program data. That is, actual student portfolios will be assessed independently by two different faculty members. A second type of data will consist of faculty judgments of a number of illustrative cases. Within each institution several independent faculty judgments will be made for each case. The illustrative cases will be adapted from actual portfolios but will be altered so as to ensure anonymity. Within a given institution estimates of the credit to be awarded for each of the illustrative portfolios will be made by independent readings of the materials.

Requirements: Institutions must be willing to provide duplicate independent assessments of credit to be awarded for at least 20 actual local students. The independent assessments will be made using the local mode of assessment. Each participating institution will also provide a minimum of five independent assessments of from 15 to 20 illustrative cases.

Stage 1:
Content
Validity

Study #B2 - Do faculty agree on what constitutes college-level learning?

Rationale: Students submit a variety of types of experience as evidence of college-level learning. A basic requirement for the validity of crediting practices is that faculty agree on which types of learning are creditable.

Data: Judgments by faculty of whether each of a series of approximately 25 illustrative examples of learning are creditable as college-level learning.

Requirements: Field research sites must be willing to provide cooperation of a minimum of five faculty to independently judge the set of illustrative cases.

Stage 3:
Concurrent
Validity

Study #B3 - To what extent do faculty agree on the type of documentation required as evidence of experiential learning, and to what extent are various types of documentation authentic?

Rationale: Students often seek credit for experiences which should and can be substantiated with documentation. The extent to which documentation is requested and the specific kinds of documentation requested may vary, however. This study would examine the degree of consistency among faculty with respect to the amount and nature of documentation required for illustrative examples of experiential learning. Also, because of limited resources it is often not possible for thorough checks to be made on the authenticity of all documentation. It is possible to make more thorough checks for a limited number of cases for research purposes however. If this is done over a sample of institutions the authenticity of documentation in general, and in various categories, can be estimated.

Data: A set of 25 illustrative cases will be presented to faculty. Each faculty member will be asked to indicate whether more documentation is needed and, if so, what type of documentation is needed. To examine authenticity documentation of various types in actual student portfolios will be checked for authenticity (e.g., a letter of reference from a former employer would be checked by phoning the employer for verification).

Requirements: Field research sites must be willing to provide the cooperation of a minimum of five faculty to judge the cases. Each field site will also be asked to provide verification data on the documentation in at least 25 student portfolios.

All Stages:
Educative
Value of
Assessment

Study #B4 - What is the educational value to the student in developing and having a portfolio assessed?

Rationale: If portfolio development and assessment are of educative value then a significant gain should be observed in student growth and development before and after assessment--especially with respect to understanding of personal competencies and educational goals.

Data: Material relating to student growth and development will be obtained for a sample of students prior to assessment and for another sample of students after assessment has taken place. Such material will be judged blind according to pre-determined criteria relevant to the educational objectives of the assessment process.

Requirements: Institutions must provide semi-structured free response measures of student growth for a minimum of 20 students randomly divided into those for whom measures are collected prior to assessment and those for whom measures are collected after assessment.

AREA C - WORK EXPERIENCE

Stage 5:
Reliability

Study #C1 - Do faculty agree on the amount of credit to be awarded for work experience?

Rationale: In order to be fair and valid, crediting practices must be consistent. One way of assessing the consistency of crediting practices is to see how closely assessors agree on the number of credits to be awarded after independently examining the same evidence.

Data: Two different types of data will be collected. The first type will be actual data obtained using the work assessment model (Working Paper No. 8) with actual students. A second type of data will be faculty judgments of a number of illustrative cases. The illustrative cases will be adapted from actual records but will be altered so as to ensure anonymity Several faculty within each institution will provide independent judgments of each case.

Requirements: Institutions must be willing to provide independent assessments of credit to be awarded for at least 20 students seeking credit for work experience. Each participating institution will also be required to provide a minimum of five independent assessments of from 15 to 20 illustrative cases.

Stage 2:
Construct
Validity

Study #C2 - To what extent can faculty articulate work experience with different occupational programs?

Rationale: Faculty within an institution should be able to assign specific competencies as relevant to courses in a particular degree program with a high rate of consistency.

Data: Judgments of faculty at field research sites as to which competencies are relevant to each of four selected occupational programs. The work competencies to be judged will be presented in the form of approximately 20 illustrative cases.

Requirements: Each research site must be willing to provide a complete set of faculty judgments for the illustrative cases. A minimum of five faculty per research site will be required.

Stage 3:
Concurrent
Validity

Study #C3 - What is the concurrent validity of the documentation of work competence?

Rationale: A basic element of the work assessment model (Working Paper No. 8) is the inventory of work competencies. To be valid there should be a high degree of agreement between the judgments of students and their supervisors as to which competencies were actually learned on a particular job.

Data: Inventories filled out by students and supervisors will constitute the basic data for the study.

Requirements: Field research sites must be willing to provide completed inventories for at least 20 students and their supervisors.

Stage 5:
Placement
Validity

Study #C4 - What is the validity of advanced placement in occupational programs on the basis of work experience?

Rationale: Students are often placed in advanced course sequences on the basis of prior work experience. This study will seek to determine whether students so exempted are as generally competent as those who have gone through the relevant course. The study will also follow such students through subsequent coursework, and compare their performance, and perceptions with traditional students.

Data: End of advanced course performance measures and student responses to a questionnaire concerning the appropriateness of the course for their background will constitute the data for this study.

Requirements: Field research sites must be willing to provide end of course performance records and questionnaire responses for at least 10 traditional and 10 non-traditionally placed students in an advanced course related to an occupational program.

AREA D

Study #D1 - Demonstration Validation of Expert Judgment.

Unlike the studies previously described, in this area individual institutions will carry out a model study that demonstrates how an institution might go about establishing the validity and reliability of its own assessment procedures. CAEL will outline a generalizable model study of expert judgment that can be applied to free response material, interviews, products, or performance. Institutions electing this option will carry out such a demonstration study following the model provided. CAEL will disseminate selected reports of such studies prepared by institutions as illustrations of exemplary quality control of assessment procedures.

The participating institution will follow a set of general principles and steps designed to result in an objective evaluation. Although the specific content may differ from study-to-study the following basic steps should consistently apply:

1. Establish crtierion standards which have a clear meaning for the assessor and student.
2. Devise an appropriate assessment procedure.
3. Conduct a pretest and revision of the assessment procedure.
4. Establish special conditions for the collection of data for the demonstration validity study.
5. Use the procedure to conduct assessments for at least 20 students.
6. Summarize the data according to described procedures.
7. Write a final report detailing the development of the assessment procedure and giving the results of the data analyses. A common format for the final report will be suggested.

The outcome for the institution will be a locally validated assessment procedure which can be used to meet future assessment needs. These local studies can be quite valuable to the institution as a means of directing faculty attention to the requirements of sound assessment procedures. The more general outcome for CAEL will be a series of brief reports demonstrating how institutions can establish the reliability and validity of expert judgment in a variety of contexts by a variety of techniques.

Schedule for Institutions Participating in Field Research

1975

July	1. Develop detailed plans for institutional tryout of experimental assessment materials at different levels of involvement (i.e., formal review, experimental use, research site)
	2. Organize institutional participation in such experimental tryout
	3. Disseminate condensed reports of all special projects completed during 1974-75
	4. Ad hoc external committee selects up to five special projects for further development and experimental tryout during 1975-76 (in addition to the seven developmental tasks)
Aug-Sept	5. Develop detailed plans for reliability and construct validity studies
	6. Distribute instruments and materials for tryout of experimental assessment materials
	7. Disseminate working papers based upon selected special projects
Oct	8. Hold workshops at Assembly meeting to discuss selected special projects in detail
	9. Hold second round workshops at Assembly meeting to discuss detailed plans for tryout of experimental assessment materials and validation studies at research sites

Oct 75 - March 76	10.	Institutional tryout of assessment materials from developmental tasks and special projects
1976		
Jan	11.	Workshop for representatives from research sites
Mar	12.	Feedback from all institutions regarding their experience with the experimental assessment materials (student and faculty reaction, research data, local applications and adaptations)
Apr	13.	Collation and analysis of information from experimental tryout of materials
May	14.	Assembly discussion of feedback and tentative conclusions regarding revision and additions to assessment materials
May-July	15.	Preparation of working papers on validity and reliability work
June	16.	Selection of institutional adaptations for publication
June-Sept	17.	Revision of all assessment materials
Sept-Dec	18.	Publication and dissemination of all revised assessment materials
Oct	19.	Presentation and discussion of institutional adaptations at Assembly meetings. Discussion of results of validity and reliability studies
Nov-Dec	20.	Preparation of final report on validation phase of CAEL project

Conditions of Participation in Field Research

A. Any CAEL member intending to experiment with CAEL materials in one or more of the four developmental areas has the further option of applying as a field research site for 1975-76. During the summer of 1975 an external committee will select a limited number of applicants that seem best suited to the research needs of the project. Such institutions will be designated in CAEL literature as Field Research Sites.

B. Applicants may designate one or more developmental areas for participation as a field research site, but one of those designated must be A, B, or C. In choosing an area the institution agrees to participate in all of the studies in that particular area.

C. Willingness to participate as a research site implies a willingness to share data on a confidential basis. Anonymity of individuals and institutions will be protected.

D. Willingness to participate as a CAEL research site implies an institutional commitment of staff resources to participate in the field research locally, to endorse such participation for students and faculty, and to send a representative to Assembly meetings in order to enhance the mutual benefit of such participation.

E. Institutions participating as field research sites will receive $2,000 from CAEL as partial reimbursement of miscellaneous local expenses and clerical costs of gathering data.

F. CAEL will provide gratis copies of relevant Working Papers for research purposes (up to 10 copies of faculty guides and 50 copies of student guides).

Criteria for Selecting Field Research Sites

An external committee appointed by the CAEL Steering Committee will select a limited number of field research sites. Institutions represented on the Steering Committee qualify as a field research site if they meet the conditions of participation. Other applicants (up to a total of 30) will be selected on the following criteria:

1. Conditions of participation must be met.

2. A diversity of institutions will be sought with respect to geography and type.

3. Institutions must be able to provide adequate data that is relevant to the studies in the area of interest indicated.

4. Reasonable balance should be maintained regarding the number of institutions participating in Areas A, B, C, and D (at least six in each area if possible.

5. Evidence of institutional commitment to the work.

Applying as a Field Research Site

Institutions applying as field research sites should complete the attached form and return it to ETS no later than July 15 with a brief description of the institution and relevant programs plus any information considered pertinent to the institution's qualifications as a research site.

Application for Participation as a CAEL Research Site

This institution would like to participate as a CAEL Research Site during 1975-76. We understand the conditions of participation and have indicated below the area(s) in which we would like to contribute data for validation studies. We will make every effort to provide data for all of the studies included in those areas we have marked below.

		Areas in which we will participate (mark with X)	Likely number of participants available*	
			Faculty	Students
A	Assessing Interpersonal Skills			
B	Use of Portfolios to Assess Non-Sponsored Learning			
C	Assessing Learning Outcomes of Work Experience			
D	Use of Expert Judgment in Assessing Learning Outcomes			

*Estimates are needed so that materials can be provided.

Institutional Representative

Name of Institution

Address

Telephone

Return by July 15 to: CAEL
Educational Testing Service
Princeton, New Jersey 08540
Attention: Jane Porter

APPENDIX 2c

SCHEDULE FOR CAEL VALIDATION WORK

1975	
March-July	1. Develop detailed plans for tryout of experimental assessment materials at different levels of involvement
	2. Organize institutional participation in such experimental tryout
	3. Ad hoc external committee selects research sites and special projects for further development and experimental tryout during 1975-76
Aug.-Sept.	4. Develop detailed plans for reliability and construct validity studies
	5. Distribute specimen instruments and materials for tryout of experimental assessment materials
	6. Disseminate working papers based upon selected special projects
October	7. Hold workshop at Assembly meeting to discuss research projects in detail
Oct. 75 - Mar. 76	8. Institutional tryout of assessment materials (working papers and institutional reports)
1976	
January	9. Workshop for representatives from research sites
March	10. Feedback from all institutions regarding their experience with the experimental assessment materials (student and faculty reaction, research data, formal models of local application and adaptation)
April	11. Collation and analysis of information from experimental tryout of materials; analyses of research data
May	12. Assembly discussion of feedback and critical issues regarding revision of working papers
May	13. Tentative results to participants and authors of working papers
June	14. Feedback to staff concerning comments on the tentative results
June	15. Selection of institutional reliability studies for publication; Executive Committee review of revised assessment reports

1976 (Cont'd)

August	16. Review of first draft of validation report by participants and authors
September	17. Review of validation report by Executive Committee
October	18. Discussion of validation report at Assembly meeting
Sept.-Dec.	19. Publication and dissemination of revised assessment materials
Nov.-Dec.	20. Preparation of final report on validation phase of CAEL project

cael
COOPERATIVE ASSESSMENT
OF EXPERIENTIAL LEARNING

A JOINT PROJECT OF EDUCATIONAL TESTING SERVICE AND A GROUP OF COLLEGES AND UNIVERSITIES WITH INITIAL FUNDING BY CARNEGIE CORPORATION OF NEW YORK AND SUPPLEMENTAL SUPPORT FROM THE FORD FOUNDATION, THE FUND FOR THE IMPROVEMENT OF POSTSECONDARY EDUCATION AND LILLY ENDOWMENT, INC.

ETS, Princeton, New Jersey 08540 ■ 609-921-9000

MEMO TO: PARTICIPANTS IN CAEL LEVEL 1 OR 2 TRYOUT.

FROM: WARREN W. WILLINGHAM

Thank you for your interest in participating in the tryout of CAEL materials. We have enclosed the faculty and/or student guides you requested, provided your request did not exceed the established maximum (5 faculty guides and 10 student guides for Level 1; 10 faculty guides and 25 student guides for Level 2). If you wish to purchase additional copies from CAEL you may do so. Please use the enclosed publications order form.

Sometime in early 1976 you will receive a questionnaire on the adequacy of the materials you are reviewing. We will look forward to receiving your completed questionnaire, as well as your further critical review and recommendations, by next March.

If you are trying out the materials at Level 2, and if you develop an "institutional model" (adaptation of the materials for use on your own campus), we will look forward to receiving that report by March 15. All such reports received will be considered for publication or distribution by CAEL.

If you have any questions during the course of the tryout, please contact Jane Porter at ETS, 609-921-9000, ext. 2186.

Prepared July 25, 1975

cael
COOPERATIVE ASSESSMENT
OF EXPERIENTIAL LEARNING

A JOINT PROJECT OF EDUCATIONAL TESTING SERVICE AND A GROUP OF COLLEGES AND UNIVERSITIES WITH INITIAL FUNDING BY CARNEGIE CORPORATION OF NEW YORK AND SUPPLEMENTAL SUPPORT FROM THE FORD FOUNDATION, THE FUND FOR THE IMPROVEMENT OF POSTSECONDARY EDUCATION AND LILLY ENDOWMENT, INC.

ETS, Princeton, New Jersey 08540 ■ 609-921-9000

August 6, 1975

The CAEL External Committee for the Selection of Participating Institutions met on July 30-31 to consider applications for participation in Level 3 tryout activities. I am pleased to inform you that your institution has been selected as a CAEL field research site in the area(s) you designated.

You will receive from CAEL a grant of $2,000 ($1,000 in September and $1,000 in February) to help defray local expenses connected with this research. You will recall that the conditions of participation are described on pp. 9-10 of the field research description you received with the application form.

Under separate cover we are sending you gratis copies of the working paper(s) for your area in the numbers indicated on your application (up to a limit of 10 copies of faculty guides and 50 copies of student guides). This quantity should suffice for research and tryout purposes. The CAEL Steering Committee expects institutions to purchase any working papers they may wish to use beyond this number.

In early September we will be sending you additional materials including a schedule of activities and instructions for collecting the necessary data for the studies. As you may already know, the CAEL Assembly meeting will be held in Philadelphia October 1-3 at the Bellevue Stratford Hotel. One of the sessions at the meeting will be devoted specifically to field research. It is an important meeting and I look forward to seeing you there. In the meantime if you have questions please contact Dick Reilly in this office (609-921-9000 Ext. 2794).

For your information, a list of all the institutions selected for the Faculty Development Program is enclosed. Congratulations on your selection. I hope that your participation will prove rewarding and productive.

Sincerely,

Warren W. Willingham

WWW/ls

Enclosures (2)

cael

A JOINT PROJECT OF EDUCATIONAL TESTING SERVICE AND A GROUP OF COLLEGES
AND UNIVERSITIES WITH INITIAL FUNDING BY CARNEGIE CORPORATION OF NEW YORK
AND SUPPLEMENTAL SUPPORT FROM THE FORD FOUNDATION, THE FUND FOR THE
IMPROVEMENT OF POSTSECONDARY EDUCATION AND LILLY ENDOWMENT, INC.

ETS, Princeton, New Jersey 08540 ■ 609-921-9000

Memorandum for: PARTICIPANTS IN CAEL LEVEL 1 OR 2 TRYOUT

From: Warren W. Willingham

Date: December 8, 1975

This memorandum comes as a reminder that we are very interested in getting your reactions to the CAEL materials you are reviewing. Sometime in early 1976 all institutions participating at Level 1 or 2 will receive a questionnaire soliciting comments from you and your colleagues who are familiar with the reports.

For Level 2 Participants

As you know, if you are using one or more of the CAEL Working Papers experimentally (Level 2), you have the option of developing an institutional model or adaptation, which will be considered for publication as part of the revised version of the report. May I remind you that the deadline for such adaptations is March 15.

The idea of encouraging adaptations by institutions stemmed from CAEL's interest in advancing sound assessment principles while avoiding undesirable standardization of procedures. The intention is not to urge institutions to merely edit CAEL reports for local use but to develop and report alternatives that are consistent with good assessment principles and represent a different approach to a particular problem, especially a different approach that fits local circumstances or objectives. Consequently, CAEL is especially interested in receiving written descriptions of local assessment procedures (i.e., adaptations) that:

a) focus upon a specific aspect of assessment related to one of the 12 reports being reviewed;

b) describe a useful variation from procedures outlined in the CAEL report;

c) give a rationale for the variation in terms of different local circumstances, different program objectives, improved efficiency and effectiveness, etc.;

d) are limited to about 10 double-spaced pages (20 pages at the outside, including any illustrations or attachments).

These institutional adaptations are to be submitted by March 15 and reviewed during the next four weeks by authors of reports to which they refer. All adaptations will be reviewed by the Executive Committee in early May with advice from the earlier review. Selected adaptations will be edited and included in a final section of the appropriate report with proper designation of author and institution. We will let you know in May whether your adaptation is selected. If the Executive Committee recommends revisions, your revised adaptation will need to be completed by June 15.

Memorandum for: CAEL FIELD RESEARCH SITE COORDINATORS

cc: D. Horch J. Porter
M. Keeton R. Reilly
J. Knapp A. Sharon

Subject: Several items for your information

Date: November 20, 1975

From: Warren W. Willingham

1. You have received or will soon receive the materials for validation studies in Area A, B, C, or D depending upon which area you are signed up for. The bulk of each packet consists largely of multiple forms and illustrative material for participants, but, of course, some institutions are working on more than one area and have received an especially heavy load. I am pleased to report that the Executive Committee has endorsed a plan for giving some recognition to differential participation by institutions. Any institution participating in more than one of the four areas will receive a supplementary allocation of $500 per extra area (beyond the first) to help offset costs of participation. You will receive that supplement in February when the second half of the original payment of $2000 is scheduled. You are to use these lump payments in whatever way you feel best supports the field research activity. I urge you to think ahead since no other funds will be available for this purpose other than travel expense to the January meeting.

2. For operational as well as fiscal reasons, it is important now to clarify definitely in which areas each institution will be working. The attached list (Exhibit 1) includes a couple of additional institutions added in Area C at the suggestion of the external committee and some changes based upon conversations with you. It represents our current understanding of who is participating in what. Please check the list and indicate on the return form if it is accurate.

3. The January 14 meeting will be held at the Hyatt-Regency--O'Hare in Chicago. In the near future we will send you a hotel registration form and ask that you make your own reservations. Please plan to pay your own bill; we will distribute travel vouchers and reimburse you later. The meeting will start with a plenary session at 8:30 a.m. and break into three sessions on Areas A, B, and C respectively for the remainder of the morning. The afternoon will be devoted largely to individual consultations. As indicated in my November 4 memorandum, it would be very useful if you can have your institution represented in each of those three concurrent morning sessions in which you are working. CAEL will pay your travel expenses plus those of any additional representative necessitated by your working in more than one of Areas A, B, and C. Please return the attached form by December 5 indicating who will attend from your institution. At the meeting Dick Reilly, Joan Knapp, and Amiel Sharon will coordinate discussion in those three areas. Over the next several weeks, one of them will be in touch with you by telephone to discuss your progress and any questions or problems that may have come up.

4. A couple of people have asked for clarification concerning how the working papers are used in relation to the validation work. You will be using them basically in two ways. First, we very much want to get some feedback from you in the early spring concerning your experimental use of the materials. After the first of the year we will be sending you a few "feedback forms" in order to solicit reactions from faculty and students who have worked with the materials. Second, the working papers will be used in support of the 12 validation studies, though all of the studies do not necessarily require students or faculty to use the working papers directly. The use of the working papers in relation to the studies tends to vary because the nature of the studies varies. Some studies test *assumptions* that are implied by the assessment procedures outlined in the working papers (e.g., the assumption that faculty members can agree reasonably well in making certain types of judgments). That type of study is typically based upon simulated materials not upon use of the working paper itself. The second type of study tests assessment *practices* and is more likely to involve student and faculty use of relevant working papers. Exhibit 2 of this memo outlines briefly the use of working papers in relation to the 12 studies. You have received the working papers you requested at the time you made application for field research. Please check over Exhibit 2 and if you feel that your needs are different than those perceived earlier, please discuss it with Jane Porter or one of the three staff members mentioned in paragraph 3 above.

5. The more lengthy enclosure is concerned with the "Review, revision, and production of CAEL reports during 1976." This document includes a variety of topics. The topics that may be of most interest to you are the following:

1) The section on the "validation report" explains how the studies in Areas A, B, and C will be reported, reviewed, and published as a CAEL report.

2) The section on "adaptations" describes a procedure whereby all institutions using CAEL materials experimentally can write up a modification of an assessment procedure for possible inclusion in one of the revised Working Papers and Institutional Reports.

3) The section called "Review Process" describes the procedure through which your Area D manuscript will be reviewed, revised, and reported.

You have already received suggestions concerning how the Area D studies might be carried out. For further information concerning how your manuscript is to be handled see Attachments 1, 5, and 6.

6. If you have questions over the next 2 months:

Call Jane Porter for information concerning meetings, communications, reports, editorial matters, etc.
Call Dick Reilly about technical questions concerning collection of data (or Joan Knapp for Area B or Amiel Sharon for Area C)

Please route this memorandum to colleagues as you feel appropriate. We look forward to seeing you in Chicago in January. In the meantime, do not hesitate to be in touch with us as questions arise.

Exhibit 2

Outlines of the Use of CAEL Working Papers in Relation to the 12 Validation Studies

A1 - Faculty agreement on the relationship of interpersonal skills to degree programs (validates underlying assumptions concerning assessment) The research materials incorporate material from Working Paper No. 4 and Working Paper No. 5. The working papers themselves provide an additional reference to help participants understand what is meant by interpersonal skills.

A2 - Consistency of judges in assessing interpersonal skills (validates assessment practices) The research materials incorporate material from Working Paper No. 4 and Working Paper No. 5. The working papers themselves provide an additional reference to help faculty and students understand what is meant by interpersonal skills.

A3 - Educational value of assessing interpersonal skills (validates assessment practices) The institution is encouraged to use Working Paper No. 4 and Working Paper No. 5 experimentally with faculty and students.

B1 - Agreement on the amount of credit awarded portfolios (validates assessment practices and assumptions underlying assessment) The institution is encouraged to use Working Paper No. 6 and Working Paper No. 7 experimentally with faculty and students.

B2 - Faculty agreement on what constitutes college-level learning (validates underlying assumptions concerning assessment) Working Paper No. 6 and Working Paper No. 7 provide background.

B3 - Documentation: Authenticity and agreement on what is required (validates assessment practices and assumptions underlying assessment) Working Paper No. 6 and Working Paper No. 7 provide background.

B4 - Educational value of portfolio assessment (validates assessment practices) The institution is encouraged to use Working Paper No. 6 and Working Paper No. 7 experimentally with faculty and students.

C1 - Agreement on the amount of credit awarded for work experience (validates assessment practices and assumptions underlying assessment) The institution is encouraged to use an assessment procedure similar to that described in Working Paper No. 8.

C2 - Faculty agreement on the relationship of work experience to degree programs (validates underlying assumptions concerning assessment) Working Paper No. 8 provides background.

C3 - Dependability of student reports of work experience (validates assessment practices) Research materials are based upon material in Working Paper No. 8.

C4 - Validity of advanced placement on the basis of work experience (validates assessment practices) The institution is encouraged to use an assessment procedure similar to that described in Working Paper No. 8. (This study was eliminated after the January meeting.)

D1 - Demonstration validity studies (validates assessment practices) The procedures are based upon material from Working Paper No. 10. The nature of the assessment procedure should be described in the Institutional Report of the study.

cael
COOPERATIVE ASSESSMENT
OF EXPERIENTIAL LEARNING

A JOINT PROJECT OF EDUCATIONAL TESTING SERVICE AND A GROUP OF COLLEGES AND UNIVERSITIES WITH INITIAL FUNDING BY CARNEGIE CORPORATION OF NEW YORK AND SUPPLEMENTAL SUPPORT FROM THE FORD FOUNDATION, THE FUND FOR THE IMPROVEMENT OF POSTSECONDARY EDUCATION AND LILLY ENDOWMENT, INC.

ETS, Princeton, New Jersey 08540 ■ 609-921-9000

Morris T. Keeton, Chairperson
Provost and Vice President
Antioch College

Richard J. Allen
Director, Division of Arts & Sciences
The Johns Hopkins University Evening College

George Ayers
Vice President
Metropolitan State University

Barbara A. Barbato
Director of the Contract Center for
Individualized Studies
Webster College

Neal R. Berte
Vice President for Educational Development
and Dean—New College
University of Alabama

James D. Brown, Jr.
President
Thomas A. Edison College

Arthur W. Chickering
Vice President for Policy Analysis and Evaluation
Empire State College

Hortense Dixon
Vice President for Urban Programming
Texas Southern University

John S. Duley
Director, Field Experience Program
Justin Morrill College
Michigan State University

Sheila Gordon
Associate Dean for Cooperative Education
LaGuardia Community College

Cyril O. Houle
Professor of Education
University of Chicago

Winton H. Manning
Vice President, Development
Educational Testing Service

Jules O. Pagano, Vice Chairperson
Dean, Community Affairs Area
Florida International University

Jean M. Pennington
Director, Continuing Education for Women
Washington University

David H. Provost
State University Dean
New Program Development and Evaluation
California State University and Colleges

Gilberto de los Santos
Dean of Instructional Development
El Paso Community College

Peter P. Smith, Secretary
President
Community College of Vermont

William G. Thomas
Chancellor
Johnston College
University of Redlands

■

Warren W. Willingham
Executive Director for Program Research
Educational Testing Service
Project Director

Memorandum for: PEOPLE ATTENDING JANUARY 14 CAEL FIELD RESEARCH MEETING

Subject: Agenda

Date: January 7, 1976

From: Jane Porter

We look forward to seeing you in Chicago next week. Although the meeting will not begin formally until Wednesday morning at 8:30, there will be an informal social hour from 5:30 to 7:00 on Tuesday evening, January 13, in the Diplomat room on the L2 Level at the Hyatt-Regency-O'Hare. We hope your travel plans will allow you to join us then. (The Hyatt-Regency provides complimentary limousine service every 15 minutes from the airport.)

The schedule for Wednesday, January 14, will be as follows:

8:30-9:00 Meet together in the Olympic room on the International Level for introductory remarks.

9:00-12:00 Break into groups according to Areas A, B, and C for discussion of the specific research studies. Arrangements are being made for individuals from different institutions to describe how they are conducting the studies and to lead the subsequent discussion. PLEASE REMEMBER TO BRING A SAMPLE FROM YOUR DATA COLLECTION.
A - Viscount room (L2 Level)
B - Ambassador room (L2 Level)
C - Consular room (L2 Level)

10:00-10:15 Coffee break in the pool area outside the meeting rooms. I will have sign-up sheets available for afternoon consultations with Dick Reilly for Area A, Joan Knapp for Area B, and Amiel Sharon for Area C. We are planning on 30-minute time slots. You may find from conversations Tuesday evening and during the coffee break that a joint consultation with one or two people from other institutions would be beneficial. If so, you are encouraged to sign up together.

12:00-1:30 Lunch—on your own

1:30-2:30 Area D meeting in the Viscount room. Since Area D is the least structured area of field research we can use this time to discuss any problems or questions you might have. This will also give us a chance to learn exactly what you are doing and how things are going.

1:30-5:30 Individual or small group consultations available on sign-up basis:

Area B - Joan Knapp - Ambassador room.

Area C - Amiel Sharon - Consular room.

2:30-5:30 Individual or small group consultations available on sign-up basis:

Area A - Dick Reilly - Viscount room.

I will be outside the meeting rooms to handle schedule questions.

JP:gm

APPENDIX 2g

cael
COOPERATIVE ASSESSMENT
OF EXPERIENTIAL LEARNING

A JOINT PROJECT OF EDUCATIONAL TESTING SERVICE AND A GROUP OF COLLEGES AND UNIVERSITIES WITH INITIAL FUNDING BY CARNEGIE CORPORATION OF NEW YORK AND SUPPLEMENTAL SUPPORT FROM THE FORD FOUNDATION, THE FUND FOR THE IMPROVEMENT OF POSTSECONDARY EDUCATION AND LILLY ENDOWMENT, INC.

ETS, Princeton, New Jersey 08540 ■ 609-921-9000

Morris T. Keeton, **Chairperson**
Provost and Vice President
Antioch College

Richard J. Allen
Director, Division of Arts & Sciences
The Johns Hopkins University Evening College

George Ayers
Vice President
Metropolitan State University

Barbara A. Barbato
Director of the Contract Center for
Individualized Studies
Webster College

Neal R. Berte
Vice President for Educational Development
and Dean—New College
University of Alabama

James D. Brown, Jr
President
Thomas A. Edison College

Arthur W. Chickering
Vice President for Policy Analysis and Evaluation
Empire State College

Hortense Dixon
Vice President for Urban Programming
Texas Southern University

John S. Duley
Director, Field Experience Program
Justin Morrill College
Michigan State University

Sheila Gordon
Associate Dean for Cooperative Education
LaGuardia Community College

Cyril O. Houle
Professor of Education
University of Chicago

Winton H. Manning
Vice President, Development
Educational Testing Service

Jules O. Pagano, **Vice Chairperson**
Dean, Community Affairs Area
Florida International University

Jean M. Pennington
Director, Continuing Education for Women
Washington University

David H. Provost
State University Dean
New Program Development and Evaluation
California State University and Colleges

Gilberto de los Santos
Dean of Instructional Development
El Paso Community College

Peter B. Smith, **Secretary**
President
Community College of Vermont

William G. Thomas
Chancellor
Johnston College
University of Redlands

■

Warren W. Willingham
Executive Director for Program Research
Educational Testing Service
Project Director

Memorandum for: AUTHORS OF CAEL WORKING PAPERS AND INSTITUTIONAL REPORTS

Subject: Feedback Forms

Date: February 6, 1976

From: Warren Willingham and Jane Porter

Attached is a copy of the faculty feedback form for your Working Paper or Institutional Report and a student form if appropriate. Five copies of each form were sent to the attached list, which was compiled from the names you gave us and the lists of Level 1 and 2 tryout participants and field research site coordinators (Level 3). In addition, all forms were sent to the Steering Committee and participants in the Faculty Development Program and the Operational Models Project. We have asked for returns by March 15. We will send copies of all returns to the authors coordinating the revision of the report and the corresponding session at the spring Assembly Meeting. An additional copy of those forms returned will be sent to any author not at the same location as the coordinating author.

The Willingham memorandum of November 20, 1975, on revision of CAEL Working Papers and Institutional Reports, provides detailed information about where to go from here. Please review this memo carefully; we will be happy to provide additional copies upon request. As explained in the memo, your first concern when you receive the forms will be the identification of "critical issues" to be discussed at the spring Assembly meeting, which will be held May 9-12 in Denver. You will need to consult (by phone and mail) with the other authors of your report (or, in the case of Working Papers 6 and 7, with the author of the companion report) to identify 5 to 10 such critical issues. Please pool your judgments and send us a list by May 1 at the latest. Meanwhile, we will undertake a content analysis of the questions on the feedback forms that are common across reports. Insofar as possible, we will provide you with results in time for your consideration of critical issues.

Please contact Jane Porter (609-921-9000, ext. 2186) if you have questions.

gm

cael
**COOPERATIVE ASSESSMENT
OF EXPERIENTIAL LEARNING**

A JOINT PROJECT OF EDUCATIONAL TESTING SERVICE AND A GROUP OF COLLEGES AND UNIVERSITIES WITH INITIAL FUNDING BY CARNEGIE CORPORATION OF NEW YORK AND SUPPLEMENTAL SUPPORT FROM THE FORD FOUNDATION, THE FUND FOR THE IMPROVEMENT OF POSTSECONDARY EDUCATION AND LILLY ENDOWMENT, INC.

ETS, Princeton, New Jersey 08540 ■ 609–921–9000

Memorandum for: PARTICIPANTS IN THE TRYOUT OF CAEL MATERIALS
(LEVELS 1 AND 2 AND FIELD RESEARCH)

cc: Ms. Knapp
Mr. Reilly
Mr. Sharon

Subject: Feedback Forms

Date: February 6, 1976

From: Warren Willingham and Jane Porter

As you know, we are eager to find out your reactions to the CAEL Working Papers and/or Institutional Reports. We have enclosed five copies of a brief faculty questionnaire for each report you are reviewing and five copies of a student questionnaire where appropriate. (For Field Research site coordinators, we have enclosed forms for the Working Papers in the areas in which you are doing the research studies; i.e., A, B, C, or D.) Please distribute the forms to people at your institution who are familiar with the reports and ask them to complete and return them by March 15.

If your institution is involved in more than one CAEL activity, it is possible that we will ourselves have sent forms to some of the people you might contact in this regard. In such a case, please try to find others on your campus who have reviewed the reports and can complete the feedback forms.

For Level 2 Participants: If you have developed an institutional model or adaptation of one of the Working Papers or Institutional Reports, as described in the Willingham memorandum of December 1 (or December 8 for some of you), please remember that we must receive it by March 15 if it is to be reviewed by the author and considered for inclusion in the revision of the report. Because we will be forwarding copies to the authors and later to the CAEL Executive Committee, PLEASE SEND US 10 COPIES OF YOUR ADAPTATION. Each copy should be clearly identified as to institution, person or persons responsible, and to which CAEL report the adaptation refers.

gm

Morris T. Keeton, **Chairperson**
Provost and Vice President
Antioch College

Richard J. Allen
Director, Division of Arts & Sciences
The Johns Hopkins University Evening College

George Ayers
Vice President
Metropolitan State University

Barbara A. Barbato
Director of the Contract Center for
Individualized Studies
Webster College

Neal R. Berte
Vice President for Educational Development
and Dean—New College
University of Alabama

James D. Brown, Jr
President
Thomas A. Edison College

Arthur W. Chickering
Vice President for Policy Analysis and Evaluation
Empire State College

Hortense Dixon
Vice President for Urban Programming
Texas Southern University

John S. Duley
Director, Field Experience Program
Justin Morrill College
Michigan State University

Sheila Gordon
Associate Dean for Cooperative Education
LaGuardia Community College

Cyril O. Houle
Professor of Education
University of Chicago

Winton H. Manning
Vice President, Development
Educational Testing Service

Jules O. Pagano, **Vice Chairperson**
Dean, Community Affairs Area
Florida International University

Jean M. Pennington
Director, Continuing Education for Women
Washington University

David H. Provost
State University Dean
New Program Development and Evaluation
California State University and Colleges

Gilberto de los Santos
Dean of Instructional Development
El Paso Community College

Peter B. Smith, **Secretary**
President
Community College of Vermont

William G. Thomas
Chancellor
Johnston College
University of Redlands

■

Warren W. Willingham
Executive Director for Program Research
Educational Testing Service
Project Director

Sample

Faculty Feedback Form

CAEL WORKING PAPER NO. 7

Because you have had some experience with CAEL Working Paper No. 7, A Student Handbook on Preparing a Portfolio for the Assessment of Prior Experiential Learning, we would appreciate learning your reactions to it. Since we will be copying the returns to share with the author, please type or write in ink. (The first four questions are merely to help us understand the perspective from which you view the report. If it is impossible to answer in terms of your entire institution, please substitute "department" or other unit for "institution.") Use additional sheets if necessary.

1. This institution is: _____2-year _____4-year _____public _____private

2. Which of the following characterize(s) your institution's interest in CAEL? (Check one or more.)

   _____We are a fairly traditional institution interested especially in giving students credit for learning acquired elsewhere but comparable to our coursework.

   _____We are especially interested in improving assessment of off-campus field experience and internship learning.

   _____We are often regarded as innovative. We are especially interested in advancing different approaches to learning and encouraging learning outcomes different from those of conventional programs.

3. This institution's experience and interest in these activities is: _____recent _____at least several years old

4. The program in which we used (or could use) Working Paper No. 7 is: _____institution wide

   _____a special program or office

5. Have you made an adaptation of Working Paper No. 7? _____If so, please send us a copy of the complete report as you have adapted it. (If you are a Level 2 participant, you may also be submitting a 10- to 20-page version of your adaptation. In that case, please send the complete adaptation in addition, not instead.)

   _____A complete copy is enclosed. _____A complete copy is being forwarded under separate cover.

   _____I am a Level 2 participant and will submit a 10- to 20-page adaptation to Warren Willingham by March 15.

6. Do you find the criteria for awarding credit, found on page 3 of the Working Paper, useful? Which criteria would you delete? What criteria would you add? Please suggest any re-wording of the criteria which you would find useful.

7. Which of the approaches to determining credit, listed on pages 36-40, would be most appropriate to your institution? Briefly describe alternative approaches that might be useful.

8. What would you identify as particular strengths or weaknesses in this report?

9. Can you cite specific issues concerning assessment practice or policy that bear on the revision of this report and might be profitably discussed at the spring CAEL Assembly meeting?

10. Everything considered, do you feel that this report is useful or not useful in your particular institution (or educational unit)? Do you have any suggestions for improvement?

Please return by March 15 to: Jane Porter, CAEL, Educational Testing Service, Princeton, NJ 08540. This form will be shared with the author of the report. Your signature is optional.

Name____________________________ Institution____________________________

Student Feedback Forms
For
CAEL WORKING PAPER NO. 7
A STUDENT HANDBOOK ON PREPARING A PORTFOLIO FOR THE ASSESSMENT OF PRIOR EXPERIENTIAL LEARNING

The CAEL project is a research and development project working on better methods of assessing student learning outside the classroom. Your college is one of the institutions participating in this project, and you have had an opportunity to work with or at least examine carefully one of the experimental CAEL student guides. The purpose of this form is to get your reactions to the guide. We will greatly appreciate your help by answering the following questions in whatever way you feel is appropriate. Please type or write in ink and use additional sheets if you need more space.

1. Please indicate how you used the guide.

   A. Check one:

   _____ I read the guide but had no occasion to actually use it.

   _____ I used the guide in connection with a course or learning activity.

   B. Check one:

   _____ I read or used the guide pretty much on my own.

   _____ A faculty member or job supervisor was actually involved in my use of the guide.

2. Do you find the criteria for awarding credit, found on page 3 of the Working Paper, useful? Which criteria would you delete? What criteria would you add? Please suggest any rewording of the criteria which you would find useful.

3. Which of the approaches to determining credit, listed on pages 36-40, would be most appropriate to your institution? Briefly describe alternative approaches that might be useful.

4. What struck you as particular strengths or weaknesses in the guide?

5. Everything considered, do you feel that this guide was useful or not useful for you in your particular learning situation? Do you have suggestions for improving the guide?

Please return this form to whoever asked you to complete it. All forms should be sent by March 15 to: Jane Por
CAEL, Educational Testing Service, Princeton, NJ 08540. This form will be shared with the author of the report
Your signature is optional.

Name______________________________ Institution__________________________

TO: CAEL FIELD RESEARCH PARTICIPANTS, AUTHORS OF WORKING PAPERS AND INSTITUTIONAL REPORTS (See Distribution List)

FROM: Warren W. Willingham

DATE: May 28, 1976

SUBJECT: Tentative results of CAEL Field Research

Thanks to the cooperation and hard work of the institutional participants, we are making good progress with the CAEL Field Research. As promised, we are now sharing with you a brief report of tentative results (plus a copy of the Field Research Progress Report, which most of you saw in draft). We had hoped to have the tentative results to you by May 15, but most of the data were late coming to us, some as much as five weeks, and we were not able to make up all of that. The overall schedule is still intact, however:

June 1 -- Area D reports due at ETS (to be reviewed by staff and authors of Working Paper No. 10)

June 18 -- feedback to staff concerning tentative results of Areas A, B, and C

June 29 -- Executive Committee selects Area D reports; review of feedback from tentative results

August 13-27 -- external review of full report

September 15 -- review of validation report by the Executive Committee

November -- publication of validation report

Most institutions were able to obtain the necessary assistance from faculty and students, and most of the data collected arrived here in good shape. But it is usually the case with complex field research of this sort that inevitable mishaps and unanticipated difficulties reduce the amount of usable data. We have had those problems, but I would emphasize that, on the whole, the data collection has been successful, and the data should be quite adequate to meet practically all objectives of the research. The data reported in the enclosures includes most of what we have, though a few returns will be added later and undoubtedly some additional analyses and refinements will be undertaken.

Let me point out what is and is not here. In addition to the progress report, we are enclosing reports of tentative results for the 10 multi-institutional studies in Areas A, B, and C (the Area D institutional reports due June 1 will be reviewed separately and those selected by the Executive Committee will appear in the validation report). Since these tentative results are intended to be a preliminary view of the findings for those already familiar with the studies, the enclosures do not include descriptions of the procedures and methods of analysis that will be part of the full report. You can see from the attached outline that the validation report and its appendix will provide descriptions of the studies in considerable detail.

Please understand that these tentative results are based upon a preliminary analysis, and the data have not yet been double-checked. When everything is rechecked, it is entirely likely that an error here or there will be corrected. Furthermore, we have not yet had time to consider the results carefully nor obtained input from you and others close to the work. Thus, I would appreciate you sharing these tentative results only with those who have been involved in the work. At this early stage of analysis, we have purposely avoided much interpretation but will be seeking reviews of draft chapters for the full report (see outline attached) in mid-August.

Naturally, participating institutions will be especially interested in their own data. In the blanks below I have filled in (on the site supervisor's copy only) the code letter that refers to your institution in the tables showing results for Areas A, B, and C.

Institutional Code

Area A ________

Area B ________

Area C ________

You will recognize that in many instances the averages or coefficients referring to a single institution are based upon a limited amount of data and are, therefore, likely to fluctuate widely simply because of normal sampling error. In the full report we will discuss the matter of data interpretation in detail and caution against placing too much emphasis on data from a single institution. In the meantime, it is probably not wise to assume that a particular statistic in a particular table is necessarily characteristic of your institution. Remember that the main purpose of this research is to test assumptions and principles; the data of individual institutions are not, by themselves, sufficiently stable or representative to permit program evaluation. A main reason for displaying data of individual institutions is to show how much the data vary across institutions so that institutions can better judge the possible significance of the departure of their own from that of institutions generally.

Please go over the reports in the area(s) of most interest to you. Authors should take these results into account in revising Working Papers and Institutional Reports but should assume that the "validation report" is the primary report through which these data will be described, interpreted, reviewed, and so on. The plan is to include in the revision of Working Papers Nos. 4, 6, 8, and 10 a brief section on "validity and reliability of assessment." This section ought to make clear the importance of gathering information to monitor and test assumptions about the assessment process. That section ought also to refer in narrative to the main outcomes of relevant aspects of field research and refer to the validation report for extended discussions and details of the field research. Other authors need not include any discussion of these results, but you may wish to refer to some specific finding and cite the validation report as appropriate.

All recipients, especially field research participants, are invited to send comments and suggestions concerning implications of these data, how they are presented, and any extenuating circumstances or qualifications that should be noted. Site supervisors have already offered such comments, but the actual results may well elicit additional suggestions or cautions. If you have comments, please send them to me (by June 18), and I will share them with the appropriate people. We want to be sure to take into account all relevant feedback in writing the complete report of this research and in reviewing other CAEL publications.

/lmc
enclosures

cc: CAEL Steering Committee

WORKING PAPERS AND INSTITUTIONAL REPORTS (1975)

Working Papers

Paul Breen, Thomas Donlon, Urban Whitaker	CAEL Working Paper No. 4 -- The Learning and Assessment of Interpersonal Skills: Guidelines for Administrators and Faculty
Paul Breen, Thomas Donlon, Urban Whitaker	CAEL Working Paper No. 5 -- The Learning and Assessment of Interpersonal Skills: Guidelines for Students
Joan Knapp	CAEL Working Paper No. 6 -- A Guide for Assessing Prior Experience Through Portfolios
Aubrey Forrest	CAEL Working Paper No. 7 -- A Student Handbook for Preparing a Portfolio for the Assessment of Prior Experiential Learning
Amiel Sharon	CAEL Working Paper No. 8 -- A Task-Based Model for Assessing Work Experience
Hadley Nesbitt	CAEL Working Paper No. 9 -- A Student Guide to Learning Through College-Sponsored Work Experience
Richard Reilly, et al.	CAEL Working Paper No. 10 -- The Use of Expert Judgment in the Assessment of Experiential Learning

Institutional Reports

Antioch College (Robert Lewis)	Institutional Report No. 1 -- The Use of Expert Judgment in the Assessment of Demonstrated Learning in the Antioch College Yellow Springs Adult Degree Completion Program
Empire State College (Lois Lamdin, Miriam Tatzel)	Institutional Report No. 2 -- Interpersonal Learning in an Academic Setting: Theory and Practice
University of California at Los Angeles (Jane Permaul, et al.)	Institutional Report No. 3 -- Evaluation and Expert Judgment

Institutional Reports (Continued)

Wayne State University (Kurt Spitzer, Sue Smock)	Institutional Report No. 4 -- The Refinement and Modification of an Instrument for Assessing the Achievement of Interpersonal Skills of Social Work Students
William Rainey Harper College (Frank Christensen)	Institutional Report No. 5 -- Guidelines and Procedures for the Assessment of Experiential Learning and for the Selection of Training of Field Experts

REVISED REPORTS (1976)

Guides and Handbooks

Breen, Donlon, Whitaker	Teaching and Assessing Interpersonal Competence--A CAEL Handbook
Breen, Donlon, Whitaker	Learning and Assessing Interpersonal Competence--A CAEL Student Guide
Knapp	Assessing Prior Learning--A CAEL Handbook
Forrest	Assessing Prior Learning--A CAEL Student Guide
Sharon	Assessing Occupational Competencies--A CAEL Handbook
Duley, Gordon*	College-Sponsored Experiential Learning--A CAEL Handbook
Nesbitt	College-Sponsored Experiential Learning--A CAEL Student Guide
Reilly, Churchill, Fletcher, Miller, Pendergrass, Porter Stutz, Clark	Expert Assessment of Experiential Learning--A CAEL Handbook

Institutional Reports

Empire State College	Institutional Report No. 2 -- Prospects and Methods for Interpersonal Studies
University of California at Los Angeles	Institutional Report No. 3 -- Documentation and Evaluation of Sponsored Experiential Learning

[The other three Institutional Reports were retained in their original form.]

---

*This report was added.

APPENDIX 3

Results of Experimental Tryout

3a. Critical Issues

3b. Content Analysis of Evaluation Forms

Tuesday, May 11
8:30-9:30

CRITICAL ISSUES: SESSION 1

THE ASSESSMENT OF INTERPERSONAL SKILLS
(CAEL Working Papers 4 and 5)

The following issues will be discussed at this session:

1. Assessment techniques and instruments

   - Would it be desirable to give greater emphasis to the interviewing technique and provide more examples of instruments?

   - What other specific instruments should we try to develop and present?

2. Standards

   - Is it important for us to make more specific suggestions about how to decide what is and isn't college-creditable? Or should we continue to consider that this is primarily a matter of local decision?

   - Should we include more discussion (and specific proposals) about criterion standards?

3. Theory

   - What uses should we assume (and encourage) can be made of the theoretical chapters in Working Papers 4 and 5?

   - Should we replace the theory chapter in the student guide with a summary--perhaps in the Appendix?

Tuesday, May 11
8:30-9:30

CRITICAL ISSUES: SESSION 2

ASSESSING PRIOR LEARNING VIA PORTFOLIOS
(CAEL Working Papers 6 and 7)

The following issues will be discussed at this session:

1. Should the two guides attempt to further define or take a position on what constitutes "college-level" learning or should they simply continue to suggest alternative approaches to faculty and students? What approaches should be cited? (Some suggestions on alternative wording will be presented at the session.)

2. Should the two guides imply that students be required to show a relationship between "old" learning and proposed "new" learning or should this criterion for awarding credit be presented as merely desirable? Some colleges use this as an important criterion, others argue that all "college-level" learning should be eligible for credit.

3. Should the two guides imply that documentation of learning activities (i.e., experience) be a required step or should this step be presented as optional? Some college argue that we are interested only in assessing learning outcomes and not verification of how learning was acquired.

4. Should transcript development be included as a step in the portfolio process as described in Working Paper 6? Alternatives in reporting non-sponsored experiential learning could be presented and pros and cons for each alternative could be discussed. Faculty members might become more effective advocates and/or decision-makers regarding the issues involved in transcripting.

Tuesday, May 11
8:30-9:30

CRITICAL ISSUES: SESSION 3

MODEL FOR ASSESSING WORK EXPERIENCE
(CAEL Working Paper No. 8)

The following issues will be discussed at this session:

1. Specific versus general competencies.
2. Need to relate work experience to course outcomes.
3. Need for the seven steps in the Work Assessment Model.
4. What is credit-worthy work experience?
5. Cost of assessment.
6. Support and involvement of faculty in the assessment process.
7. Appropriateness of the norm-referenced approach for establishing standards of competence.
8. Relevance of the Work Assessment Model to the Liberal Arts Curriculum.

Tuesday, May 11
8:30-9:30

CRITICAL ISSUES: SESSION 4

LEARNING THROUGH COLLEGE-SPONSORED WORK EXPERIENCE
(CAEL Working Paper No. 9)

1. To how broad an audience should the student guide be addressed? One faculty respondent has determined that the current version of the guide has a "fog level" of 14.2, placing its vocabulary level at about the sophomore year of college. Some community college faculty have suggested that the guide is too sophisticated in its present form for their students to use without continuous instruction or supervision. On the other hand, several readers have argued that the publication is of little appeal to graduate students or mature undergraduates.

How much revision can be accomplished to give the guide a broader appeal without limiting its usefulness to any single audience? While the same chronological approach can be used in any case, considerations of the intended audience would affect the style of writing, the choice of examples used for illustration, and the amount and type of self-directed activity prescribed. (Note: The companion faculty guide that is currently in preparation should to some extent lessen the need for the student guide to serve a much broader audience.)

2. How can the format and graphics be improved to make the guide more useful and attractive? Readers who have been quite complimentary about the text of the working paper have on the other hand expressed concern that the present graphic layout makes it visually unappealing--at least to students. (We are considering reasonably inexpensive improvements, including dual columns, illustrations, and more blank space per page.) Also, a number of respondents have indicated that they would prefer to use just certain portions of the guide (e.g., forms and checklists). This could be accomplished by perforating the pages and punching holes in them for insertion in loose-leaf notebooks.

3. What specific content areas need to be improved in the guide? The main suggestions thus far have called for amplifying the section on assessment and giving the topic of assessment more prominence in the guide; for giving more detailed explanations of some of the suggested procedures (i.e., performance analysis, values clarification); and for making some of the examples more realistic. Are there others?

Tuesday, May 11
8:30-9:30

CRITICAL ISSUES: SESSION 5

EXPERT JUDGMENT IN THE ASSESSMENT OF EXPERIENTIAL LEARNING
(CAEL Working Paper No. 10)

The following issues will be discussed at this session:

1. Monitoring the quality of individual assessment: How does one deal with the problems of reliability and validity when only a few people are involved?

2. To what extent can non-structured as opposed to structured interviews be used and how does that influence outcomes? What is the role of the expert in deciding what type of interview to use?

3. What kinds of procedures should be used in developing criterion standards and how should accompanying rating scales be developed? What is the role of the expert in this process?

4. In the chapter on Product Assessment, should the section on "General Considerations" be expanded? Should it include a more detailed description of how to conduct a product assessment?

5. The validity-reliability dilemma in performance testing: How does one deal with the problem that increasing the validity of a performance test may decrease its reliability and vice versa?

6. How can the difference between student assertion and actual learning outcomes be clarified? How can we explain the need for an assessor to sort descriptions of learning opportunities from actual college-level learning deserving of credit?

Tuesday, May 11
8:30-9:30

CRITICAL ISSUES: SESSION 6

ASSESSING THE DEMONSTRATED LEARNING OF ADULTS
(Institutional Report No. 1: Antioch College)

The follwoing issues will be discussed at this session:

1. Expectations and reality level of student preparation for assessment.
2. Faculty participation in assessment--for example:
    - Relationship of assessment to other work that faculty do
    - Payment or non-payment
3. Timing of evaluation--before or after admissions?

The people attending this session will divide into three groups, each focusing on one of the above issues.

Tuesday, May 11
8:30-9:30

CRITICAL ISSUES: SESSION 7

INTERPERSONAL LEARNING IN AN ACADEMIC SETTING
(Institutional Report No. 2: Empire State College)

The following issues will be discussed at this session:

1. The definition and operation of interpersonal learning in varied educational institutions.

   Recognizing that there are a range of traditional and innovative programs that are or are not involved in interpersonal learning, how can Empire State College's experiences assist them in beginning and advancing student work in this area?

2. How to structure learning activities to develop interpersonal competencies.

   Based on our observation that attempts to develop interpersonal competencies at Empire State tend to be implicit or tangential to the student's learnings rather than explicit and direct, it is critical to devise procedures for deliberately attempting to foster interpersonal learning.

3. How to measure changes in interpersonal competence.

   Although many existing contracts address learning and growth, few attempt to assess where a student stands at the outset, the degree of change, and how to make those determinations.

Tuesday, May 11
8:30-9:30

CRITICAL ISSUES: SESSION 8

EVALUATION AND EXPERT JUDGMENT IN COLLEGE-SPONSORED EXPERIENTIAL LEARNING
(Institutional Report No. 3: University of California, Los Angeles)

1. In refining CAEL Institutional Report No. 3 and the utility of its content, we have been working under the premise that the Record-Reflection Log and Matrix should be the vehicles for the following:

    - An instrument which encourages and enhances student self-assessment and provides, through the assessment process, enlightenment on what students have actually learned through a given unit of experience.

    - A documentation of activities, the learning derived from them, the level of competence, and, indirectly, the learning process.

    - A source of information essential for formative and summative evaluation of course and program.

    - A source of information for determination of grades and creditation.

    The first critical issue is whether these premises are the ones with which we should be most concerned. Are there others? Are some of these of lesser importance?

2. The adequacy of the report in orienting students and log analysts in the appropriate use of the Log and Matrix--is special orientation or training essential? What should be the key elements to be emphasized either in the Report or at orientation training?

3. Value of Matrix information--is the information which appears on the Matrix adequate for the purposes of evaluation of either the student or program? Should we suggest learning categories and proficiency levels? Should there be more specific definitions? What about the frequency of log-keeping--daily, weekly, monthly?

4. The Log analysis process--what are the economics and paper work involved? What about the qualifications of the analyst? Are important pieces of information from the Logs lost in the abstracting process?

5. Validity of student-report learning as reflected by the student--is student judgment about his/her learning acceptable?

Tuesday, May 11
8:30-9:30

CRITICAL ISSUES: SESSION 9

AN INSTRUMENT FOR ASSESSING INTERPERSONAL SKILLS
(Institutional Report No. 4: Wayne State University)

The following issues will be discussed at this session:

1. How to use an instrument focusing on specific competencies and utilizing rating scales, in a humanistic individualizing manner.

2. How to bring together the service philosophy of people responsible for the field training with the philosophy of the educational institution in order to provide the student with an integrated framework for experiential learning.

3. How to develop a variety of modalities to assist field teachers in effective teaching and meaningful assessment of experiential learning.

4. How to develop useful methods for evaluation of the effectiveness of the various modalities used in the preparation for teaching and the assessment of experiential learning of field teachers.

Tuesday, May 11
8:30-9:30

CRITICAL ISSUES: SESSION 10

DEVELOPING INSTITUTIONAL GUIDELINES FOR THE ASSESSMENT OF EXPERIENTIAL LEARNING

(Institutional Report No. 5: William Rainey Harper College)

The institutional guidelines or tentative procedures described in Institutional Report No. 5 include a section on portfolio development. The critical issues discussed at this session will focus on a student workbook for portfolio preparation. A brief tape/slide show will be presented.

Table 1. Percentage of Respondents Who Cited Different Kinds of Strengths

	REPORT						
	A %	B %	C %	D %	E %	F %	G %
Style (well-written, clear, etc.)	12	8	8	8	15	5	7
Organization, format	17	10	8	3	12	10	9
Length (brevity)	0	0	0	3	3	0	5
Comprehensiveness	4	13	10	42	9	12	16
Level	6	3	3	0	1	0	0
Methodology, procedures	10	0	8	6	4	3	14
Theory, ideas, concepts, models	6	10	13	6	15	2	23
Practicality, applicability	10	3	10	3	9	0	5
Examples	2	10	3	3	3	12	9
Graphics, forms, checklists	13	0	0	0	9	2	7
Particular passages or sections	21	23	33	19	38	10	16
Generally positive comment	25	15	5	14	18	17	11
Other strengths	4	5	3	3	6	0	7
Total number of forms	52	39	39	36	78	58	43

Table 2. Percentage of Respondents Who Cited Different Kinds of Weaknesses

	REPORT						
	A %	B %	C %	D %	E %	F %	G %
Style (confusing, not well written, etc.)	2	3	3	0	9	3	7
Organization, format	0	3	0	0	3	3	16
Length, redundancy	19	3	5	14	12	3	14
Level too difficult	2	5	8	6	21	10	7
Too theoretical, too hypothetical, theory too complex	0	0	5	0	9	0	7
Not enough theory	0	0	3	0	1	0	2
Not practical, not applicable, methodology as weakness	10	8	10	11	12	12	16
Examples (inappropriate, not applicable)	0	5	0	0	0	0	0
Examples (too brief, not enough)	0	5	3	6	6	3	2
Graphics, forms, checklists (confusing, irrelevant, dull, difficult)	8	0	0	0	9	2	0
Graphics, forms, checklists (too brief, not enough)	2	0	0	0	1	2	0
Particular passages or sections	6	13	18	3	27	3	26
Generally negative comment	4	0	0	0	1	0	0
Other weaknesses	10	13	8	14	8	3	9
Total number of forms	52	39	39	36	78	58	43

Table 3. Percentage of Respondents Who Made Different Kinds of Suggestions for Improvement

	REPORT						
	A %	B %	C %	D %	E %	F %	G %
Clarify; edit (reference to language)	0	0	0	0	6	0	0
Structure differently; edit (reference to structure)	0	0	0	0	1	0	5
Make more concise, shorter; eliminate redundancy; edit (reference to length)	8	0	0	3	4	3	0
Make more specific for a given audience	0	3	0	3	8	7	12
Abstract; summarize	0	0	3	3	3	3	2
Add examples	2	8	3	6	9	0	5
Add graphics (figures, tables, charts)	0	0	3	0	0	0	0
Add forms, worksheets, models, sample instruments	2	0	0	0	1	0	5
Modify graphics	0	3	0	0	8	0	2
Add a section	4	5	3	3	3	2	0
Rewrite or change a section	4	10	8	6	10	0	7
Add more theory	0	0	0	3	0	0	2
Add more how-to-do-it, practical material	0	3	3	3	4	0	0
Other suggested improvements	4	5	0	8	12	9	5
Total number of forms	52	39	39	36	78	58	43

Table 4. Percentage of Respondents Who Cited Different Kinds of Issues

	REPORT						
	A %	B %	C %	D %	E %	F %	G %
ISSUES RELATED TO:							
Identifying learning outcomes	0	5	3	11	0	3	12
Articulating learning outcomes to educational goals	2	3	5	0	0	0	2
Documenting learning activities, outcomes	0	0	3	3	1	0	0
Measuring the extent and character of learning outcomes	0	10	18	14	6	2	7
Judging the learning outcomes	0	0	8	6	3	2	2
Evaluating whether the learning outcome meets acceptable standard	6	5	10	11	3	2	5
Transcribing the credit or recognition	0	0	3	0	0	3	0
Facilitating assessment	4	8	15	14	0	2	2
Cost/benefit issues	0	3	5	8	0	2	0
Educative value of assessment	0	0	0	0	3	0	0
Other (not codable in this framework)	17	8	15	6	13	5	28
Total number of forms	52	39	39	36	78	58	43

APPENDIX 4

Research Materials

Area A - Interpersonal Skills

4a. Guidelines for Site Coordinators
4b. General Faculty Guidelines for Interpersonal Skills CAEL Validation
4c. Illustrative Cases for Study A1: Management, Education, and Psychology
4d. Faculty Rating Forms for Study A1: Management, Education, and Psychology
4e. Structured Interview Guidelines for Study A2
4f. Role Playing Exercise for Study A2
4g. Faculty Rating Form for Study A2
4h. Student Questionnaire for Study A3

Area B - Portfolio Assessment

4i. Guidelines for Site Coordinators
4j. Illustrative Portfolio for Study B1
4k. Faculty Portfolio Evaluation Report for Study B1-Illustrative Portfolios
4l. Institutional Portfolio Summary Sheet for Study B1-Actual Local Portfolios
4m. Fifteen Illustrative Examples of Prior Learning Outcomes for Study B2
4n. Faculty Recording Form for Study B2
4o. Fifteen Illustrative Examples of Prior Learning Outcomes for Study B3
4p. Checklist for Recording Type of Documentation Required for Study B3
4q. Form for Recording Verification of Documentation for Study B3 (Guidelines for Determining Document Authenticity for Study B3)
4r. Student Questionnaire for Study B4
4s. Site Coordinator Questionnaire for Study B4

Area C - Work Experience

4t. Guidelines for Site Coordinators
4u. Work Experience Evaluation Form for Study C1: Data Processing, Law Enforcement, Secretarial Science, and Management
4v. Illustrative Case for Study C2
4w. Faculty Rating Form for Study C2: Data Processing, Law Enforcement, and Management
4x. Student Checklist for Study C3: Law Enforcement (with Faculty Instructions)

Area D - Demonstration Validity Studies

4y. Guidelines and Procedures

Revised Guidelines for Site Coordinators

Area A - Interpersonal Skills

The purpose of these guidelines is to provide research site coordinators with as much detail as possible concerning the background of the validation studies and the steps which should be taken to ensure consistency in experimental procedures and data collection. First, a brief description of the studies is given, then a series of steps is presented for each study which should be approximately followed at each field research site. Finally, an overall schedule for the validation phase is given.

Description of the Studies

Study #A1 - To what extent can interpersonal competencies be consistently articulated with different degree programs?

The basic purpose of this study is to examine the extent to which faculty agree on the relevance of various types of interpersonal skills to different programs of study. The three areas of study are management, education, and psychology. A total of 15 simulated student portfolios will be presented to each of three faculty judges within each of three areas of study (a total of nine judges). Each faculty judge will be asked to read through each of the 5 student portfolios in his or her area and then to rate the potential credibility of the student's learning for a given area of study. The data provided by faculty at your institution along with data from other institutions will then be analyzed so that conclusions can be drawn regarding the extent of agreement among faculty.

Study #A2 - To what extent do different assessors agree on the level of interpersonal competence students have acquired and to what extent is there agreement between different methods of assessment?

The purpose of this study is to examine the extent to which there is agreement between two different methods of assessing the same set of competencies or skills and between two different assessors within each of these methods. The substantive area chosen for assessment is that of counseling. For purposes of this study counseling will include any kind of one-to-one helping relationship. The first method which will be used to assess counseling competence is a structured interview. The second method will be role playing simulation in which the student is cast in the role of a counselor and the assessor is cast in the role of a student. Two different teams of evaluators will be needed for this experiment. The first team will conduct the interviews with each assessor alternating as interviewer and observer. The second team will conduct the role playing assessments again with each of the two assessors alternating as participant and observer. Two assessors for each technique are needed because not only will the agreement between the interview and role playing assessment methods be studied but also the agreement within each method will be studied by examining the independent assessments made by the two assessors within each team. A minimum total of 15 students should be assessed by both methods. The structured interview provided can be regarded as a general instrument which may be modified to fit your particular needs.

Study #A3 - What is the educational value to the student of the assessment of interpersonal skills?

The purpose of this study is to examine the effects of the assessment process on student personal growth and development especially as it relates to the educational goals of the student. Data will be needed from a total of 20 students. For 10 of the students the data will consist of responses to a questionnaire which will be administered prior to assessment. The remaining 10 students will provide data in the form of responses to the same questionnaire administered after assessment has taken place. Responses to the questions will later be rated by a panel of CAEL faculty members. Comparisons will then be made between overall quality of the pretest group versus the posttest group responses. Institutions will be asked to utilize local adaptations of the material in CAEL Working Papers Nos. 4 and 5 in conducting their assessments and in preparing students for assessment. For each of the 10 students for whom post-assessment data is collected the site coordinator will be asked to provide information describing the assessment process.

Steps in the Collection of Data

1. Check Materials Needed

All materials necessary for completing data collection have been provided. As a first step, however, the site coordinators should become familiar with the materials and make sure that an adequate supply of each item is available. Materials which will be supplied are as follows:

Study A1

- Ten (10) copies of General Faculty Guidelines
- Four (4) sets of 5 simulated cases in management
- Four (4) sets of 5 simulated cases in education
- Four (4) sets of 5 simulated cases in psychology
- Ten (10) copies of faculty rating forms

Study A2

- Four (4) copies of General Faculty Guidelines
- Four (4) copies of the structured interview guidelines for faculty
- Four (4) sets of faculty guidelines for the role playing method
- One hundred (100) copies of student rating scales

Study A3

- Four (4) copies of General Faculty Guidelines
- Thirty (30) copies of student questionnaires
- Fifteen (15) copies of assessment description form

2. Plan Data Collection

a. Faculty cooperation: For Studies A1 and A2 faculty cooperation will be required. For Study A1 a minimum of three faculty will be required to judge the illustrative cases in each of the three areas. For Study A2 two teams of two faculty members each must agree to conduct the interview and role playing assessments. The agreement of participants should be obtained as early as possible. There are no specific criteria for the selection of faculty participants. However, you are urged to select faculty who are representative of those normally involved with the assessment of experiential learning. Information on faculty, such as a profile, would be helpful but is not necessary and, in fact, if faculty wish to remain anonymous they may do so by using a code number rather than their name.

b. Obtain student cooperation and schedule student participants: For Studies A2 and A3 student participants will be involved. It will be necessary to obtain student cooperation and to schedule students for each of the two assessment methods in Study A2 and to coordinate the student schedules with the assessor schedules. For Study A3 it will be necessary to arrange for the pretesting of half of the students to be assessed and the posttesting of the other half. For the minimum target of 20 students the procedures should be as follows: First, assign a number from 1 to 20 to each of the 20 students. This number should match the chronological sequence in which the students are to be assessed. For all of the odd numbered students the questionnaire should be administered prior to assessment. For all of the even numbered students the questionnaire should be administered after assessment has taken place.

3. Meet with Cooperating Faculty

Two separate meetings should be held before data collection begins. The first meeting should include those faculty members who will be involved in rating the simulated cases. The purpose of the meeting will be to acquaint faculty with the research studies, to answer any questions that are raised and to give specific details on how data are to be recorded and collected. One of the points which should be emphasized at this meeting is that the ratings or judgments must be made independently. That is, there should be no collaboration or discussion of the ratings made by different faculty members until after all ratings have been completed and returned to the site coordinator. Participants should also be informed at this time as to arrangements for collecting their completed ratings. A second meeting should be held with faculty involved in Study A2. The purpose of the study should be explained and the part which each faculty member will play should be clarified. It should be pointed out to faculty members that they are to alternate roles from participant to observer in the assessment method with every other student. All faculty members involved in Study A2 should thoroughly familiarize themselves with the assessment materials. That is, the interview schedule, the role playing method, and the rating scales.

It will be particularly important for participants in Study A2 to maintain regular communication with the site coordinator. Also, code numbers should be assigned to all faculty participants at the time of the initial meeting.

4. Collection of Data

Before the 1st of January at least 1/3 of the data in each of the three studies should be collected. The remaining 2/3 of the data can be collected after January 1, but no later than the beginning of March.

5. Attend Meeting for Site Coordinators in Early January

A meeting will be held in early January for all site coordinators at which time any problems or difficulties which have been encountered in the collection of data can be raised and hopefully ironed out.

6. Return Data to Educational Testing Service

Data should be returned no later than March 1, 1976.

7. Schedule

The following are key dates for site coordinators in planning and executing the data collection.

Date	Event
Mid October	CAEL Assembly Meeting
Early November	Complete Plans for Data Collection
Mid November	Complete Meetings with Faculty Participants and Schedule
January 1	Complete at least 1/3 of the Data Collectio in each of the Three Studies
January 8	Attend Meeting for Site Coordinators
March 1	Finish Data Collection and Mail to ETS

General Faculty Guidelines for

Interpersonal Skills CAEL Validation

As a faculty member participating in the validation of CAEL it may be helpful for you to be provided with the definitions of interpersonal skills which were utilized in CAEL Working Papers Nos. 4 and 5.

Interpersonal skills may be defined broadly as those competencies that lead to effective relationships between an individual and one or more other persons. Defined in this way, interpersonal skills would be demonstrated by the political leader who is addressing a convention attended by thousands of people. More narrowly, the definition can be limited to those skills that affect purposive, face-to-face encounters between two persons or some group small enough to afford immediate interpersonal feedback. These guidelines do not presume any such narrowing, although the principal emphasis is on small-group settings. The guidelines are oriented toward a wide range of situations involving interpersonal relationships. In fact, interpersonal competence is seen as an adaptive process, one of selecting appropriate behaviors. The concept which is used to denote this flexible ability is "interpersonal literacy," knowing when (a) and how (b) to communicate what (c) to whom (d) in order to achieve specified goals (e).

a. WHEN to communicate (the timing of interpersonal communication). A person possessing adequate interpersonal skills will know when to interrupt and when to listen; whether to offer a comment or make a suggestion at the beginning of a conversation, or later after necessary groundwork has been laid; whether to scold (or praise) immediately after an action, or later, with more perspective or perhaps a different group of listeners, etc.

b. HOW to communicate (the medium of interpersonal communication). A person who is interpersonally literate will know whether to transmit a message verbally (and whether to do it in writing or orally) or whether (and how) to use nonverbal techniques.

c. WHAT to communicate (the content of interpersonal communication). The person skilled in interpersonal communications knows what information, questions, or feeling to transmit.

d. To WHOM to communicate (the target of interpersonal communication). The interpersonally literate person knows which audience to select and, when communicating with more than one person, whether or not to send identical messages, whether or not to inform each person that others are receiving the same (or different) messages, etc.

e. IN ORDER TO ACHIEVE SPECIFIED GOALS (the purposes of communication). The person who is interpersonally literate will be able to use communications skills to maximize the attainment of goals that are realistic and to recognize when they are not realistic.

Faculty are referred to CAEL Working Paper No. 4 for a more detailed discussion of the area of interpersonal skills.

APPENDIX 4c

STUDY A1 -- CASE M3
MANAGEMENT

DEGREE PACT COVER SHEET

NAME: Jenny DuPre

Number of Transfer Credits 41.8

I completed Business College in 1957, at 17, fully intending to apply my increased earning power toward a Degree. Like many other young people, I procrastinated--enjoying all the luxuries and rewards of full-time employment. Over the years, I changed secretarial positions several times (continually increasing responsibilities and monetary rewards) did a great deal of traveling and much inner-assessment of who I was and what I really wanted.

I married at 24, becoming the mother of a boy and a girl. I consider my family life to be my most valuable and exciting life experience. It has never occurred to me that I would not serve my family's needs (as well as my own) best through continuance of a career. Full-time employment was resumed after a three-week leave-of-absence for the birth of each child. We have a closely-knit, working-and-playing-together family, where each member is sensitive to the needs, goals and independence of each other.

I have strived throughout my entire life to improve my skills and broaden my knowledge in every way available to me. In my opinion, at this time I can be of most value to myself, my family, my employer and my community, by further development of communicative and management decision-making capabilities. When this series of studies is completed, I intend to more closely define continued-learning processes into one or two specific areas which will best heighten my career potential and value. For the timebeing, my aim is to build and broaden my experience through acquisition of general knowledge in a variety of areas.

I hope to enlarge my personal development through study of humanistic psychology and perfecting my present knowledge of interior decorating. Basic psychology should add to all aspects of work and private life.

Vocationally, I will expand into the areas of advertising, business law, marketing, financial statement analysis, contract administration, metrication, economics, management by objectives, computer program requests, reading more efficiently and creative problem-solving, all of which should further my career and make a more valuable contribution to my employer.

My community involvement will be improved with knowledge acquired in the areas pf public speaking and understanding the structures of state and local governments.

In keeping with a personal aim of increased total family involvement, I hope competence at the game of tennis will add to our list of enjoyable family pastimes.

CONSULTANTS USED IN DESIGNING THIS EDUCATIONAL PROGRAM

Name	Position or Title
Max Eberhart	President, Eastern Hardware Company
Ralph Mason	Consultant
George Anderson	Consultant

This will indicate, as of this date, I intend to implement the attached Degree Pact and I understand the procedure to be followed if I decide to modify this Pact.

Jenny DuPre

Date December 3, 1974

January 20, 1975

Personnel Supervision and Training:

Jenny DuPre knows and can apply the theories and techniques of personnel supervision and training.

She has attended several seminars on administrative supervision; was Senior Confidential Secretary in charge of Steno Pool for 5 years; employed as Secretary to the Senior partner of a law firm for 4 years; employed 2 years as Secretary to Branch Manager of an Electronic firm, with supervisory responsibilities; 18 months as Senior Secretary in Purchasing Department. She is presently employed as my Secretary. She has trained clerical people in proper office procedures, either in group sessions or on a person-to-person basis, correcting and evaluating the results; has developed instructional manuals for office procedures; and conducted training sessions to clarify procedures.

She has demonstrated to my satisfaction, a sound working knowledge of the following criteria:

(a) Able to establish rapport with fellow employees, achieving respect.

(b) Able to obtain pleasant working relationship without social involvement.

(c) Able to carry out results required by Superior, handling day-to-day work-flow by delegating assignments when necessary, in an efficient manner.

(d) Able to show interest in work-oriented problems of co-workers without getting personally involved.

(e) Able to obtain voluntary employee feedback and communication valuable to management-employee relationships, without violating confidentiality of discussions.

(f) Development and presentation of training manual for dictaphone use.

(g) The above session was clear, concise and informative.

(h) There has been considerable improvement in the general office as it relates to dictation capabilities.

(i) Employees attending the session found it and the manual to be helpful and informative.

Maxwell Anderson Eberhard
President and Chief Executive Officer
Eastern Hardware Company

EASTERN HARDWARE COMPANY

January 20, 1975

To Whom It May Concern:

As President and Chief Executive Officer of Eastern Hardware Company, I am totally and completely responsible for the overall operational and financial management of the corporation. As Chief Executive Officer, I am responsible to the Stockholders in a legal sense for all Security Exchange Commission Filings' compliance as well as legally responsible for compliance with all Federal and State Statutes affecting our corporate behavior.

Mrs. DuPre, who is Chief Administrative Secretary to my office, is obviously involved in the total corporate affairs and as such, must of necessity demonstrate a complete understanding and sensitivity to all of the ramifications of top corporate management. Not only does she carry out her normal secretarial responsibilities in a most professional manner, but additionally, because of the demands of my office, must become increasingly aware of and have knowledge of, all of the various operational functions within the corporate system. Operating on a Line and Staff concept, Mrs. DuPre must not only be totally aware of Line and Staff responsibilities but additionally, the functions within each Staff Division. She, in effect, is being exposed to, and, in fact, is being trained in expanding and broadening her knowledge of total corporate management.

As a professional corporate top management man, I have been involved in top management responsibilities for the past 15 years and prior to that time, spent 10 years in every functional area of corporate operations, ranging from finance, to control, marketing and merchandising.

Sincerely,

Maxwell Anderson Eberhart

Maxwell Anderson Eberhart
President

Attachments

Name: Mrs. Jenny DuPre (John A., Jr.)

Competence: Knows and can apply theories and techniques of personnel supervision and training.

Achieved by: Seminars on administrative supervision; employment as a secretary for various firms with increasing supervisory responsibilities; training clerical personnel in office procedures including developing instructional manuals and conducting training sessions.

Assessment Technique: Observation of job performance.

Evaluated by: Max Eberhart, President and Chief Executive Officer, Eastern Hardware Company. (Supervisor)

Evaluation Summary:

- Establishes rapport with and achieves respect from employees.
- Handles day-to-day work flow efficiently, delegating assignments when necessary.
- Can obtain voluntary employee feedback and communication.
- Training session was clear, concise, and informative.

Evaluation recorded on the transcript 6/17/75

MAE/kmw

Name: Mary Jean Lyons

Competence: Knows and can apply communication and basic instruction techniques in patient health education.

Achieved by: Establishing teaching plans for diabetic patients, identifying patient needs and learning capabilities and adjusting teaching plans; guiding and supporting teaching of nursing staff.

Assessment Technique: Observation of work performance.

Evaluated by: Jesse Allen, Director, Medical Services, Emerson Medical Center.

Evaluation Summary:

- Assessed "tone" and learning capabilities of groups and adjusted teaching program accordingly.
- Related generalities to individuals specific condition and needs.
- Can establish rapport with individuals, define her role with them, and develop teaching plan accordingly.
- Has competence in "teaching patients, individually or collectively, and their families about diabetes, its care and management."

EMERSON MEDICAL CENTER

November 1, 1973

Re: Mary Jean Lyons

Dear Sir:

Ms. Lyons has developed the competency for coordinating a study of an existing program for the purpose of modifying and improving it. The study included: conducting interviews to solicit data from various levels of nursing staff, synthetizing the collected data into a chart to demonstrate common problems and possible solutions, participating in a group analysis of the process involved in care and education of a diabetic patient.

The study also included co-writing a proposal for the restructure of that program. The proposal included: identification of the objectives and goals for a fiscal year, designing and implementing a new form to be used by members of the hospital nursing staff, identification and detailing of job responsibilities of persons involved in the care and education of a diabetic patient, preparation of her job description including: a job summary, responsibilities, relationship of herself to others in the hospital, plus recommendations for educational requirements, and an orientation program.

In implementing the program, Ms. Lyons conducted informational sessions with nursing staff members to explain the diabetic team's responsibilities and to explain the new forms and procedures. She has developed the competency for preparing a budget for that program.

She has developed the competency for preparing educational material to be used for patient education, (visual aids, brochures) and preparing material for informational purposes, (policies and procedures for various aspects of diabetes care, co-author of "Diabetes Manual," author of "Guidelines for Teaching").

Ms. Lyons has developed a competency for establishing rapport with individuals, defining her role with them, and developing a plan accordingly. With patients she identified their needs and established a teaching plan in cooperation with the physician and the nursing staff caring for the patient. She also assessed their learning capabilities to adjust the teaching plan accordingly. Her role with the nursing staff included guiding and supporting the teaching they provided, assessing their readiness, knowledge, and capabilities to provide that teaching, and providing spontaneous instruction as well as formal classes to meet their educational needs.

During her position as Diabetes Nurse Specialist, she attained the competency for teaching patients, individually or collectively, and their families about diabetes, its care and management. She assessed the "tone" and learning capabilities of the group and adjusted the teaching program accordingly; she related generalities to the individual's specific condition and needs.

Two projects for which she gained a competency were directly related to developing community educational opportunities. In 1971 and 1972, she co-developed a diabetes fair to offer lay people in the community an opportunity to learn more about diabetes, its care and management. The project included: contacting and soliciting the support of community agencies, coordinating hospital services and resources, synthetizing the above resources into a format, preparing advertisements, preparing an evaluation, compiling the data collected from the evaluations, and making suggestions for future events.

In 1973, she co-developed a diabetes workshop for community nurses to provide them with facts about diabetes pertinent to their professional role. This project included preparing a proposal for which she solicited information from the nursing groups in the community to determine their needs in diabetes education, compiled that data into a chart to illustrate the need, educational objectives and classes to meet the identified needs, and prepared a budget for each workshop. In developing the programs, she contacted community resources to assist in the program, assisted in planning of each agenda, conducted group discussions in each workshop, and compiled data collected from the evaluation.

If you have any further questions regarding Ms. Lyon's competencies, feel free to contact me at your convenience.

Sincerely,

Jesse Allen

Jesse Allen, RN MSN
Director

JA/jp

STUDY A1 -- CASE P5
PSYCHOLOGY

DEGREE PLAN COVER SHEET

Student's Name: Sue Shoemaker   Number of Transfer Credits: 80 (prov.)

In the space provided below, write a statement indicating the relationship of the specific competences in this Plan to your life goals:

The past three years have been filled with gaining understanding and love of myself. This was brought about by working with a Gestaltist and a Transactional Analysis counselor. As a result of these experiences I have developed a strong desire to understand, study and accept people. I have also developed a strong love of expressing myself and the surroundings I live in through the medias of Art and Music.

My experiences in living with other women has left a great many questions unanswered inside of me. I am wanting to gain perspective insight as to who women are. Who are the significant other's in women's lives? How do the differences between women influence their mythology, her-story, and values systems? I believe I shall gain a start in understanding these questions and many others with the program I have put together.

My overall vocational goal is to become a Human Development Counselor. Each of the competences of this pact build upon each other so a strong foundation of basic skills is developed. I intend to study more intensively several diverse forms of psychology and creative areas of expression. I have also included a few competence goals for my own needs of fun. I also anticipate a need of understanding the grief, death, and dying process. Finally, I plan on studying Sign and Spanish languages as a means of communication with a greater number of people.

Consultants used in designing this educational program:

Name	Position or Title
Raymond Turnkey	Transactional Analysis Counselors
Sally Strong	Gestaltist/co-founder of Sagaras

This will indicate, as of this date, I intend to implement the attached draft Degree Plan and I understand the procedure to be followed if I decide to modify this Plan.

Sue Shoemaker   Aug. 1975
Student's Signature   Date

The attached draft Degree Plan has been reviewed and found acceptable by the following:

Raymond Turnkey   8/75
Sally Strong   8/75

SECTION I: COMPETENCE GAINED PRIOR TO ADMISSION TO DEGREE CANDIDACY

E. PERSONAL DEVELOPMENT AND SOCIAL AWARENESS COMPETENCE

1. Competence statement:

Knows and can apply basic procedures and principles of interpersonal communications on a one-to-one non-theraputic level.

Process by which competence was gained:

Was engaged in one-to-one growth sessions with Sally Strong one of the founders of Sagaras a feminist health collective, and Raymond Turnkey a clinical trainee in Transactional Analysis for over a year.

I read *Awareness* by John Stevens, *Don't Push the River* by Barry Stevens, *Born to Win* by Muriel James and Dorothy Jongeward, and *Games People Play* by Eric Berne.

Evidence of competence:

Will be assessed by Raymond Turnkey and Sally Strong. They will use their good judgment and sense of selves to evaluate my ability.

I expect the assessments to show that:

a. I know how to ask for clarification.

b. I know how to check out assumptions.

c. I know how to work with feedback

d. I know how the readings have influenced my communication skills.

LEARNING ACTIVITY AGREEMENT

Aug. 15, 1975
date

No. 40100-7

REGISTRATION

STUDENT: Sue Shoemaker

Telephone:

☐ Degree Candidate; ☐ Non-Degree Student

Advisor _____ Lrng. Ctr. _____

SPONSOR: Sally Strong

☐ Community Faculty; ☒ Resource Person; ☐ Educ. Institution or Community Agency; ☐ Permanent Faculty

__Course/CLO; __Independent Study; __Internship; __Programmed/Correspondence Course; __Workshop/Seminar/Conferen
__Consultation; X Prior Competence Assessment; __"While-Enrolled" Competence Assessment; X Other CA

IF LEARNING ACTIVITY: Title Interpersonal Communications Pact Ref# _____

IF COMPETENCE ASSESSMENT: Pact Ref# _____ DATE(s): Aug. 1975 to _____

COMPETENCE TO BE ATTAINED/ASSESSED -- STRATEGIES, ASSESSMENT TECHNIQUES, AND CRITERIA:
[attach 3 copies from: (check one) ☐ approved Plan or ☐ Plan modification]

Assessment/Evaluation of Competence(s) attained:

Since the competence which Sue is claiming is experiential, rather than intellectual, I am basing this assessment on my own observations and experience of her over this last year.

Sue began her personal growth sessions with an intimate other, where communications were difficult and personal awareness just beginning to develop. Consequently, work was done to increase individual awareness first; so that each was clearer on what they wanted to share, and the feelings that sharing entailed. Then some real structured work on communication ensued with emphasis on checking out assumptions and dealing with feedback. She also came for many one-to-one sessions to clarify and change her perceptions and behaviors in other areas. I believe Sue has grown creatively, and has shown very clearly her abilities to clarify and check assumptions, both with her close relationships, her work and educational situations, and with me. She is continually working on her handling of feedback, particularly in staying "un-hooked" in dealing with anger which is directed at her but which is not basically connected with her. I am satisfied with her readings and especially pleased with her delight and warmth in response to Don't Push the River.

I am a gestalt therapist and have worked with women for the last year and a half. I have extensive and on-going training in my field and several years before that of related work experience. More importantly, I have worked deeply in these areas for myself and am assured of my own competence in these areas.

LEARNING ACTIVITY AGREEMENT

Aug. 15, 1975
date

No. 40100-7

REGISTRATION

STUDENT: Sue Shoemaker

Telephone:

☐ Degree Candidate; ☐ Non-Degree Student

Advisor ________ Lrng. Ctr. ________

SPONSOR: Raymond Turnkey

☐ Community Faculty; ☒ Resource Person; ☐ Educ. Institution or Community Agency; ☐ Permanent Faculty

__Course/CLO; __Independent Study; __Internship; __Programmed/Correspondence Course; __Workshop/Seminar/Conference; __Consultation; X Prior Competence Assessment; __"While-Enrolled" Competence Assessment; X Other CA

IF LEARNING ACTIVITY: Title Interpersonal Communications Pact Ref# ________

IF COMPETENCE ASSESSMENT: Pact Ref# ________ DATE(s): Aug. 1975 to ________

COMPETENCE TO BE ATTAINED/ASSESSED -- STRATEGIES, ASSESSMENT TECHNIQUES, AND CRITERIA:

[attach 3 copies from: (check one) ☐ approved Plan or ☐ Plan modification]

Assessment/Evaluation of Competence(s) Attained:

Evidence of Competence:

1. Sue has been able to ask for and accept clarification of herself and reading material. I see this ability, and knowledge in the way that she asks for and not demands for clarification. Sue has applied, "asking for clarification" in and out of groups that I have been a leader of.

2. Checking out assumption is one of Sue's biggest steps forward. Taking the risk of checking out assumption has helped me become a much closer person to Susan Shoemaker. She has applied "check out assumption" not only in group but in the outside world, in what I would call a one-to-one non-theraputic level.

3. "Working with feedback" is the only way Sue would be able to apply (1) and (2) above.

4. The readings of Awareness by John Stevens, Don't Push the River by Barry Stevens, and Born to Win by Muriel James and Dorothy Jongeward. Sue has been called upon to evaluate in group for the knowledge of other persons. She was able to identify the advantages and disadvantages of the theories. In doing this, she has been influenced in interpersonal communication.

Impression Mental Status:

She offers good eye contact, she has many facial expressions. She answers questions directly, and offers spontaneous conversation, and asked me questions. Her speech was (is) unremarkable, her memory for both recent and past events is intact. She showed no conceptual disorganization, and her thought content was not unusual. She did not describe depressive moods. She shows good judgment, ability to plan, and has been able to follow helpful directions. She is not chronically depressed, has no suicidal indication.

Impression:

Human lady in touch with herself, other persons, and the large world of interpersonal communications.

Recommendation:

Because of the quality of work I have seen Sue do in the interpersonal communications, she would be able to enroll as a student. If I can be of any other resource to her or the school, please let me know by return letter.

Aug 1975
Date

Raymond Turnkey
Signature

Faculty Rating Form for Study A1

Education

Assessor ______________________________

For each of the 5 simulated cases you are asked to judge the potential creditability of each student's learning experiences by estimating the number of credits which would be awarded in the education area. You are asked to provide separately estimates for Freshman-Sophomore, Junior-Senior and Graduate level work* as well as a total estimate. Also, feel free to provide explanatory comments regarding the reasons for your estimates.

Case #	Freshman-Sophomore	Junior-Senior	Graduate	Comments
E1				
E2				
E3				
E4				
E5				

---

*Where this breakdown is not appropriate, estimate of total credits will be sufficient.

4-21

APPENDIX 4d

Faculty Rating Form for Study A1

Psychology

Assessor ______________________________

For each of the 5 simulated cases you are asked to judge the potential creditability of each student's learning experiences by estimating the number of credits which would be awarded in the psychology area. You are asked to provide separately estimates for Freshman-Sophomore, Junior-Senior and Graduate level work* as well as a total estimate. Also, feel free to provide explanatory comments regarding the reasons for your estimates.

Case #	Freshman-Sophomore	Junior-Senior	Graduate	Comments
P1				
P2				
P3				
P4				
P5				

---

*Where this breakdown is not appropriate, estimate of total credits will be sufficient.

Faculty Rating Form for Study A1

Management

Assessor ______________________

For each of the 5 simulated cases you are asked to judge the potential creditability of each student's learning experiences by estimating the number of credits which would be awarded in the psychology area. You are asked to provide separately estimates for Freshman-Sophomore, Junior-Senior and Graduate level work* as well as a total estimate. Also, feel free to provide explanatory comments regarding the reasons for your estimates.

Case #	Freshman-Sophomore	Junior-Senior	Graduate	Comments
M1				
M2				
M3				
M4				
M5				

---

*Where this breakdown is not appropriate, estimate of total credits will be sufficient.

Structured Interview Guidelines for Study A2

Instructions for Faculty Assessors

In conducting the student interviews you are asked to use the attached outline as a general guide which may be modified or adapted for your own particular situation. For each interview only one assessor should play an active role with the other serving strictly as an observer.

At the completion of the interview the student should be thanked and excused while both assessors independently make their ratings. Please do not discuss your ratings or make any changes based on collaboration with each other.

A Structured Interview for Counseling

1. Tell me about the type of counseling you have done (setting, types of client, etc.).
2. What kinds of people have you counseled?
   a. What was the range in age?
   b. Were they both males and females?
   c. From what racial, social, and religious groups did they come?
3. How many people do you estimate that you have counseled?
4. How did you learn counseling?
5. What skills would you identify as necessary for success in your type of counseling?
6. Could you specify what personal characteristics you think an effective counselor must possess?
7. What kinds of problems have your counselees brought to you (educational, parent-child, occupational, sexual, drug-related, etc.)?
8. Could you give me a specific example of a problem which was brought to you?
9. How do you get started in a counseling session?
10. Do you have any particular way of opening up a counselee who does not seem able to tell the whole story?
11. Do long silences occur in your counseling sessions? How do you handle them?
12. How long do your sessions tend to be?
13. How do you close a counseling session?
14. How many sessions do you usually have with one counselee?
15. How do you tell if a series of counseling sessions is finished, and how do you end it?
16. To what extent do you suggest plans and activities to your counselees?
17. To what extent do you explain to a counselee what you think causes his or her problems?
18. How would you describe your personal counseling style?
19. What makes your style effective or ineffective?
20. How frequently do you refer your counselees to other sources of assistance?
   a. To what sources?
   b. For what services?
   c. How do you decide when to refer?
   d. To what extent do you follow up on your referrals?
21. How frequently have you had experience in counseling more than one person at a time?
   a. How many persons were in the typical group?
   b. What kinds of persons were in the typical group?
   c. How is your counseling approach in a group different from your approach with an individual?
   d. What is the single greatest difference between group counseling and individual counseling?

22. If you had to specify one skill which the majority of your clients lacked, what would it be?
23. Are there any consistent behavior patterns which recur from client to client that you can identify?
24. Do you feel that you perform any general counseling behaviors which are applicable to all cases or counseling situations?
25. If a client does not respond to your methods over several weeks, what do you do?
26. Has there ever been a client you could not counsel?
    a. If so, what did you do, and why?
27. If you had to identify three characteristics which make some counselees open to counseling, what would they be?
28. Do you ever do a self-assessment of your performance?
    a. How often?
29. What standards or criteria for assessing your own counseling performance do you use?
30. What criteria or frame of reference do you use to measure a client's progress?
31. Have you had your counseling assessed by any others?
    a. Were their criteria the same as yours?
32. If you had to specify one skill you need to improve or sharpen most in relation to your counseling, what would it be?
33. What kind of training do you think a new counselor in your setting should have?
34. What, if anything have you learned about yourself through your counseling experience?
35. To what extent has your own life experience contributed to your effectiveness as a counselor?
36. How do the characteristics of a counselor affect the approach to counseling that he is likely to follow?
37. How important is it for the counselor to be like the counselee in age, sex, race, etc.?
38. To what extent should counselors specialize in types of problems, like marital problems, drug-related problems, etc.?
39. What specific theories and methods do you use most often in counseling?
40. Do you follow the counseling approach of any particular author?
41. What are the goals of your job as specified by your superiors?
42. Do you think that counseling will be (remain) your life's work?
43. What is the hardest part of counseling?
44. What is the most satisfying part of counseling?

Role-Playing Exercise for Study A2

Instructions for Faculty Assessors

For this exercise one of you is asked to assume the role of a new counselee to be interviewed by the student. (The other assessor will serve as an observer.) You are asked to choose a role with which you are comfortable but at the same time one which is appropriate and relevant to the background of the student. The role you play could be taken from your own personal experience or from a case with which you are familiar.

The student will be asked to interview you for approximately one-half hour. At the end of the interview the student will be asked to summarize his perceptions of the "counselee's" problems and suggest a plan of assistance. After this has been done, the student should be thanked and excused while both assessors independently make their ratings. Please do not discuss your ratings or make any changes based on collaboration with each other.

APPENDIX 4g

Faculty Rating Form for Study A2

Please rate the student on each of the five factors listed below. Please do not discuss your ratings or change them based on collaboration. Circle the statement alongside each factor which corresponds most closely with your assessment. (Circle one number on each line.)

Skills are Equivalent to Those of a Practicing Professional (5)
Equivalent to One of More Years Graduate Training (4)
Equivalent to Two or More Years in a College-Level Precounseling Program (3)
Equivalent to One of Two College-Level Precounseling Program Courses (2)
Little or no Competence is Evident in this Area (1)

Factor					
Range of counseling experience.	1	2	3	4	5
Ability to adapt effectively and deal with unusual problems . . .	1	2	3	4	5
Knowledge of counseling methods and procedures.	1	2	3	4	5
Effectiveness as an interviewer	1	2	3	4	5
Ability to plan and carry out a program of assistance	1	2	3	4	5
Communication skills.	1	2	3	4	5
Evaluative skills	1	2	3	4	5
Overall effectiveness as a counselor.	1	2	3	4	5

Student Questionnaire for Study A3

Please answer each question as fully but as concisely as possible. Try to confine your answer to the space provided. If you need more space attach additional sheets of paper with the question numbers indicated.

1. Briefly describe what you would regard as your major strengths in the area of interpersonal skills? i.e., In what interpersonal areas are you most competent or effective?

2. Briefly describe what you would regard as your major weakness in the area of interpersonal skills? i.e., In what interpersonal areas are you least competent or effective?

3. Briefly explain how the skills you have acquired in the interpersonal area are related to your future long-range career goals?

4. Briefly explain how the skills and competencies you have acquired in the interpersonal area are related to your personal goals?

5. Briefly explain how the skills and competencies you have acquired in the interpersonal area are related to your immediate educational plans?

Revised Guidelines for Site Coordinators

Area B - Portfolio Assessment

The purpose of these guidelines is to provide research site coordinators with as much detail as possible concerning the background of the validation studies and the steps which should be taken to ensure consistency in experimental procedures and data collection. First, a brief description of the studies is given, then a series of steps is presented for each study which should be approximately followed at each field research site. Finally, an overall schedule for the validation phase is given.

Description of the Studies

Study #B1 - Do faculty agree on the amount of credit awarded for portfolios?

The basic purpose of this study is to examine the extent to which different faculty agree on the number of credits to be awarded for a student's prior learning after independently judging the same portfolio. Two different kinds of data, simulated and actual, will be used. First, a total of 14 simulated portfolios will be presented to three different teams of faculty judges each consisting of two or more faculty qualified to judge prior learning in one of the areas included in the student portfolio. Each judge will be asked to read through each portfolio which will present evidence of prior learning in one or more areas. The faculty person then will be asked to award credits in the area in which he or she is most competent to evaluate the student's learning. For example, a portfolio may present learning outcomes in business and music. A minimum of two faculty experts in each of these fields would judge the amount of credit to be awarded. In each case the music expert would award a certain number of music credits and the business expert would be asked to estimate the business credits to be awarded. These two estimates would then be added so that two independent total credit judgments would result for each simulated portfolio.

In addition to the simulated study, fifteen (15) actual institutional portfolios will be presented to faculty for duplicate credit award judgments using the local mode of assessment and evaluation. Basically this will involve an independent but complete repetition of the institutional assessment process. Portfolios will remain in the institutional files and will NOT be sent to ETS. The site coordinator will be responsible for recording the credit award on a summary sheet. The data provided by faculty at your institution along with data from other participating institutions will be analyzed so that conclusions can be drawn regarding the extent of intra- and inter-institutional agreement on the number of credits awarded for individual portfolios.

Study #B2 - Do faculty agree on what constitutes college-level learning?

The purpose of this study is to investigate the extent to which faculty agree on which prior learning outcomes and competencies are creditable in terms of higher education. Fifteen (15) illustrative examples of prior learning outcomes will be presented for judgment. Each illustration will be accompanied by the experience from which it resulted and some brief evidence of achievement. Five faculty members are needed to independently judge the set of illustrative cases. Assignment of credits to each outcome is NOT required. Data resulting from this study will be analyzed with data from other institutions to determine what types of learning are rejected or accepted by a majority of faculty as creditable.

Study #B3 - To what extent do faculty agree on the type of documentation required as evidence of experiential learning and to what extent are various types of documentation authentic?

The major purpose of this study is to determine the degree of consistency among faculty with respect to the amount and type of documentation required by them as evidence for a particular experiential-based learning. To accomplish this fifteen (15) examples of learning outcomes excerpted from student narratives will be considered independently by five faculty members. For each of the examples faculty will list their preferences for documents and/or products which would aid them in assessment. Also, because of limited resources, it is often not possible to check the authenticity of all documentation. For this reason, institutions participating in this study will be asked to provide verification data on documentation found in 25 actual student portfolios. Faculty are not required to participate in the verification portion of this study.

Study #B4 - What is the educational value to the student in developing and having a portfolio assessed?

The purpose of this investigation is to examine the effects of the portfolio process on student personal growth and development especially with respect to the understanding of personal competencies and educational goals. Data will be needed from a total of 20 students. For 10 of the students the data will result from responses to an open-ended questionnaire which will be administered prior to portfolio assessment. The remaining 10 students will provide responses to the same questionnaire administered after portfolio assessment has occurred. Responses to the questions will later be rated by a team of CAEL faculty members. Comparisons will then be made between the pretest group and posttest group in respect to the overall quality of student responses. Institutions will be asked to utilize local adaptations of the material in CAEL Working Papers Nos. 6 and 7 in conducting their assessments and in preparing students for assessment.

Steps in the Collection of Data

1. Check Materials Needed

All materials necessary for completing data collection have been provided. As a first step, however, the site coordinators should become familiar with the materials and make sure that an adequate supply of each item is available. Materials which will be supplied are as follows:

Study B1

- Five (5) copies of 14 simulated portfolios
- One hundred (100) copies of faculty instructions/portfolio recording forms
- Two (2) forms for recording faculty judgments on actual institutional portfolios

Study B2

- Ten (10) copies of a set of 15 illustrative examples of prior learning outcomes
- Ten (10) copies of faculty instructions/recording form

Study B3

- Ten (10) copies of a set of 15 illustrative examples of prior learning outcomes
- Ten (10) copies of a checklist form for recording type of documentation needed
- Four (4) copies of guidelines for authenticating documentation
- Five (5) copies of form for recording verification of document authenticity

Study B4

- Thirty (30) copies of student questionnaire
- Two (2) copies of site coordinator questionnaire about institutional components of portfolio assessment process

2. Plan Data Collection

a. Obtain faculty cooperation: For Studies B1, B2, and part of B3 faculty cooperation will be required. For Study B1 a minimum of two faculty per area of academic expertise will be needed to judge any one particular aspect of a simulated portfolio. A minimum of five faculty members will be required to judge the sets of illustrative examples in Studies B2 and B3. The agreement of participants should be obtained as early as possible. There are no specific criteria for the selection of faculty participants. However, you are urged to select faculty who are representative of those normally involved with the assessment of experiential learning. Information on faculty, such as a profile, would be helpful but is not necessary and, in fact, if faculty wish to remain anonymous they may do so by using a code number rather than their name.

b. Obtain student cooperation and schedule student participants: For Study B4 student participation is required. It will be necessary to arrange for the pretesting of half of the students who will be undergoing portfolio assessment and posttesting another half who will have completed the assessment process. For the minimum target of 20 students the procedures should be as follows: First, assign a number from 1 to 20 to each of the 20 students prior to portfolio construction and assessment. For all of the odd numbered students the questionnaire should be administered prior to portfolio assessment. For all of the even numbered students the questionnaire should be administered after assessment has taken place.

3. Meet with Cooperating Faculty

The purpose of the meeting will be to acquaint faculty with the research studies, to answer any questions that are raised and to give specific details on how data are to be recorded and collected. One of the points which should be emphasized at this meeting is that the ratings or judgments must be made independently. That is, there should be no collaboration or discussion of the ratings made by different faculty members until after all ratings have been completed and returned to the site coordinator. Participants should also be informed at this time as to arrangements for collecting their completed ratings. Also, code numbers should be assigned to all faculty participants at this time.

4. Collection of Data

Before the 1st of January at least 1/3 of the data in each of the three studies should be collected. The remaining 2/3 of the data can be collected after January 1, but no later than the beginning of March.

5. Attend Meeting for Site Coordinators in Early January

A meeting will be held in early January for all site coordinators at which time any problems or difficulties which have been encountered in the collection of data can be raised and hopefully ironed out.

6. Return Data to Educational Testing Service

Data should be returned no later than March 1, 1976.

7. Schedule

The following are key dates for site coordinators in planning and executing the data collection.

Date	Event
Mid October	CAEL Assembly Meeting
Early November	Complete Plans for Data Collection
Mid November	Complete Meetings with Faculty Participants and Schedule
January 1	Complete at least 1/3 of the Data Collection in each of the Three Studies
January 8	Attend Meeting for Site Coordinators
March 1	Finish Data Collection and Mail to ETS

STUDY B1 -- PORTFOLIO NO. 12 (JEAN CONROY)

Credits requested in: SOCIOLOGY
and/or SOCIAL WORK

Resumé

Name: Jean Conroy

Date of Birth: December 18, 1945

Department: Sociology

Major: Urban Affairs

Educational Experience:

1/68-6/68	Hyatt Institute - Leadership Training Program	non credit
	Seminar in Child Management	non credit

Employment Experience:

1/67-6/67	Hoover School (Special Ed.)	Teacher Aide
1967-1975	Big City Urban League (Operation Open City)	Housing Counselor

Community Activities:

1963-1968	Big City CORE	
1963-1971	FOCUS (Civil Rights Organization)	
1964-1968	Pilot Project for Quality Integrated Education	
1964-1971	Women's Strike for Peace	
1966-1968	Beachhaven Nursery School	Board of Directors
1968-1971	Committee for Nest Rocket Housing	
1972-	Organized consciousness raising groups in my community	

GENERAL ESSAY

For many years following the landmark 1954 Supreme Court decision determining that it was unlawful to segregate school systems, whether de facto or de jure, and with the passage of the 1958 Civil Rights act, the focus had been on the Southern campaign and the Civil Rights struggle, aided by an activist Supreme Court, was being waged in terms of social and political rights for the heretofore disenfranchised segment of U. S. society. The vitality and growing militancy of the movement energized in the South was fast spreading throughout America in a panorama of direct action: demonstrations, sit-ins, strikes, boycotts and civil disobedience. Black people were really getting it together and the call that reverberated from shore to shining shore was "Freedom Now!". Then it was August, 1963 and better than one quarter of a million Americans, responding to the call of the dashing, young knight from his White House, converged from all directions upon Washington, D. C. This was indeed the era of the great marches and this was indeed the greatest of them all. They were that day young and old, black and white, rich and poor, old-time liberals and old-line leftists, union leaders and union members; they were that day, Americans, united and exercising a great American tradition; demonstrating their belief in the need for social change. They had the blessing of the white knight, they had black and white together, they had Martin Luther King's dream, but most of all they had hope; the belief that in this country this was the way to get things done; this was the way to bring about the changes so desperately needed that would improve the condition of the black, the Puerto Rican, the poor. Most of the rest of the nation, who were not there that day, were surely if not inspired, at least neutralized. For many of those who were there, like myself, it was the first experience of many yet to come, of putting our bodies where out mouths were. It became clear during the March on Washington that black demands were shifting emphasis from equal political rights toward equal economic opportunities. President Kennedy, allegedly having been touched by Michael Harrington's "The Other America," an account of America's poor, was also fast reaching the conclusion that a "war" to end racism and provide equal opportunity for all U. S. citizens was inextricably tied to a "war" which would eliminate poverty.

Kennedy began and after his assassination, Johnson continued the legislative battle against poverty culminating in the Economic Opportunity Act, which, more than any other piece of legislation, is identified with the Great Society's commitment to our nation's poor. Its announced aim was not only to eliminate poverty but to restructure society by giving the poor a chance to design and administer antipoverty programs. Who amongst those who cared could take issue with such highsounding and sorely needed objectives? Not many....and so the placards were folded away and the Movement moved off of the streets and into the new and shiny offices that housed the spiritually and financially (for some) uplifting business of helping the poor help themselves. In the future each time the incediary conditions of ghetto life exploded, additional crumbs would be thrown the antipoverty way and may one-time-revolutionaries-turned-antipoverty workers could be found alongside the police in the paradoxical position of helping to "cool" things off. A pattern developed in which a correlation between the effectiveness of a program in "keeping the lid on" through the "long, hot summers" and the amount of antipoverty funds received could be drawn. It became clear to many that the "War on Poverty" was nothing more than the system yielding however little it could get away with, in preservation of itself. Far more continued to be expended on Johnson's real war, the one being waged against the Vietnamese people; for colonialism, whether external or internal, is the imperative of the American Capitalistic System. Still there were those who persisted in the delusion that progress was being made.

Open City, the Fair Housing arm of the Big City Urban League, began life as a pilot project in September, 1964, with a staff of three full-time workers. The program was based on the premise that equal access to housing is a fundamental right of all people and in recognition of the fact that in spite of Federal, State and City laws barring discriminatory practices in the sale or leasing of housing units, this unlawful practice continued, denying the right of equal access to Black and Puerto Rican people. In February, 1966, application was made to the city and approval received for an Antipoverty grant in the amount of $150,000 for an initial six-month period. This made possible expansion of the staff to twenty-two full-time and fifteen part-time workers and the opening of Open City offices. Approval for enlarged grants continued to be received from the Council Against Poverty and enabled Open City to provide expanded coverage by opening a fourth office. Open City operated as an autonomous unit of the Urban League, with the League's jurisdiction limited chiefly to fiscal matters.

From the beginning Open City was well-received by the communities which we served. Within five months of the first Antipoverty grant, registration more than doubled; by late 1967, more than 900 families had moved into good housing in new neighborhoods through the direct aid of Open City; more than 800 additional families had also relocated, indirectly aided through the materials and information which we distributed widely and the counseling, encouragement and general feeling of support generated by the Open City staff and successful registrants. During our initial funding period, 13% of all families served were receiving Welfare assistance (later this figure greatly increased); 65% had family incomes of less than $6,000 per year; 25% had annual incomes between $6,000 and $8,000; only 10% had annual incomes over $8,000 and most often these were joint incomes. Before the negative effects of rent decontrol were felt throughout The Big City, 30% of Open City's placements were welfare recipients or families existing on annual incomes of less than $4,000. Fifty-three percent of the placements had annual incomes of less than $6,000 and 76% less than $7,000. The correlation between racial ghettoes and poverty had been the subject of many research studies. Open City was a determined effort to help minority group families break the vicious cycle of the poverty chain by fighting the discrimination which, coupled with low income, keeps families immobilized and confined to the sub-standard housing prevalent throughout ghetto areas. Inferior housing, inferior schools, inferior consumer products and produce, inferior services, inferior everything is the blight of the ghetto, except the cost of the inferiority which is universally superior to the cost white families are paying in cummunities which systematically bar entrace to blacks and Puerto Ricans.

Although much time and effort was spent in publicising our program through the black and Spanish press, radio, clubs, churches, union, community progress centers and other antipoverty agencies, word-of-mouth remained the largest single factor in bringing new clients to Open City. Ghetto families desperate in their need for help and weary of the cold shoulder of bureaucratic red tape and indifference, came to our offices in large numbers. At first they came solely hoping to duplicate the success friends and relatives had had in obtaining apartments; later they came with any problem involving any social institution. They learned that they would be met with an honest effort on our parts to solve their seemingly and in many cases, actually, insoluble problems. Such was the policy of the agency and the commitment of the staff, that no client would be turned away because the problem for which they sought help was out of our domain or because we were not funded for that purpose. If we could not ourselves intervene on behalf of our clients directly (because we felt the client would be better served through the expertise of another agency or because it was felt a particular public course of action might put the program in jeopardy) we then made referrals to proper sources. Referrals in Open City meant more than merely filling out a slip of paper and handing it to a client. We always made sure before sending anyone, anywhere, that the service extended by the receiving agency was the most effective course of action for the client to follow; wherever possible the client was provided with the name of a contact person and an appointment made. Most of our referrals were to Welfare Rights groups, civil rights organizations, rent strike committees, volunteer attorneys and other antipoverty programs. The other type of referral which was common practice was to a Fair Housing group. Through these volunteer groups, Open City clients were escorted and introduced to predominantly white areas in which they could afford housing, but in which they had historically met with so many rebuffs that they were discouraged from looking. There was extensive follow-up on each and every referral and close contact was maintained with both the client and receiving agency until a satisfactory disposition of the case was made.

Registration was accomplished by phone, mail or an in-person interview whichever was possible for the client. The main office which was deluged with registrants, conducted daily group orientation meetings. In the Big City, we urged that clients come into the office for their interview as personal contact always resulted in the greatest service to the client. In view of the hardhip this imposed on working clients, the staff rotated turns keeping the office opened one night a week and four hours every Saturday. We provided each registrant with "Kits" of material explaining their rights under the Fair Housing Laws and specific suggestions for seeking housing on the open market. "Neighborhood Spotlights," detailed descriptions of over forty areas in all parts of the city, in terms of facilities, schools, transportation, rental range, etc., were included, along with descriptive booklets about government-aided middle-income cooperatives and rental developments. Special housing bulletins on new opportunities were issued frequently, e.g., government subsidies available under the "Capital Grant" program and the "skewed" rental plan which enables low-income families to move into middle-income developments. Information on mortgages, buying houses, loans for cooperatives, credit union loans, etc., was compiled regularly and distributed widely. All our literature was printed in both Spanish and English. Perhaps the most important booklet we issued was entitled "Know Your Housing Rights," which details the basic legal rights of recipients and discloses the amount per family unit, the Department of Social Services is authorized to pay for rent, in order that our clients learn what they were actually entitled to rather than merely relying on what they were told by their caseworker. Each interview culminated in an evaluation by a housing counselor as to the best possible direction to be given the client.

Over 60% of the moves by Open City families into new housing required filing complaints with the City Commission on Human Rights. Open City regarded this as a measure of progress in that while discrimination always existed, only now was it being effectively challenged. We informed clients of available apartments gleaned through daily scouring of the real estate sections of all newspapers (including the ethnic weeklies to which we subscribed and behind which landlords attempt to hide) and white volunteers combing through multiple dwelling units in their communities. Experience had taught us to anticipate discrimination each and every time we sent a client to check an apartment lead, therefore, careful and detailed instruction was given to the client along with all pertinent information on the apartment itself. In very short order we had the entire client/checking system down to a science and the City Commission on Human Rights recognizing our expertise authorized Open City Housing Counselors to take complaints and affadavits in their stead. We rarely lost a case, but oft times even a win could be an empty victory;

i.e., once a case went past the initial conciliatory stage without a successful resolution, a long and tedious legal battle might ensue during which time it often became impossible for the complainant to await the outcome. When this happened the usual procedure was for the apartment to be rented to another Open City applicant but this adjudication was hardly just to the original complainant. All the Fair Housing Laws lack sufficient teeth to really eliminate discriminatory practices and all the adjudicating agencies lack commitment and motivation to pursue vigorously discriminatory practices (without outside agitation) even within the limits of the existing laws (the single exception being the CCHR under the chairmanship of Robert Baxter). This became apparent when "repeating" landlords and their now "expert" attorneys proved to be the most inexorable opponents. The onus of proving discrimination is placed upon the aggrieved party, making it highly improbable for people, unaided by an Open City type operation, to obtain any satisfactory resolution of their complaints. The commission, in addition, extends to the landlord/plaintiff every possible consideration, often at the expense of the complainant; for these reasons an Open City staff person always accompanied the client to every session and/or hearing.

However, when it came to putting obstacles in the way of our welfare recipient clients the Department of Welfare took the prize. These clients have the least mobility of all and face a double-edged sword of discrimination; the racial prejudice is indemnified in this instance by the predudice directed against welfare recipients in general (this latter discriminatory practice is not yet unlawful and

Open City was working hard toward that end). The Department of Welfare jeopardized the discrimination cases involving our mutual clients by their consistent policy of non-cooperation, until, as a result of the pressure Open City brought to bear, a liaison person was appointed to deal with Open City complaints. Previous to the appointment, we had great difficulty in getting rent levels approved above the caseworkers authority and equally difficult to obtain was the issuance of rent and security checks without verification through usual Department channels. However, since those channels entailed the Department contacting the landlord and since in cases of discrimination the landlord cannot be involved before the propitious moment, subsequent to the appointment of the liaison person, Open City was permitted to be the verifying agency.

Open City also opened channels of communication in behalf of our clients with the Department of Relocation and the Big City Housing Authority and was able to make some small headway in relocating "burn-out" emergency victims and four-member family units in public housing (the greatest concentration of apartments in public housing is two bedroom apartments, therefore four-member family units constituted maximized usage). Open City's "target area" concept was focused on securing housing for minority group families in neighborhoods of high housing value and low racial mixture. The lifting of rent control on apartments had over the years made this increasingly more difficult. The inflated cost of this housing source, once our mainstay, severely curtailed the access of our clients to these apartments and in fact, severely curtailed the vacancy rate, diminishing to some degree the mobility of all residents and creating an insurmountable obstacle for the city's poor.

Open City has always received favorable treatment from funding Government agencies and, in fact, in the "Report of the White House Conference on Civil Rights," was commended as the number one action program. This acclaim in light of the low priority housing has generally been accorded in the distribution of antipoverty funds, can only be explained, in my opinion, by the fact that Open City really represented no threat to the status quo. Limited as we were to working on a one-to-one ratio, we created only little ripples and not nearly the sweeping waves of change that must occur in this society. Open City delivered important services to the black and Puerto Rican communities, but in truth we were only taking up the slack and filling the void left by already existing social service agencies and Human Rights Commissions that simply were not fulfilling their stated objectives and functions. If the Federal Government were really intent on providing good housing for the poor then why not massive Federal spending on rent subsidy programs and on large-scale construction of low-cost housing? Why not incentive programs offered to State, County and Municipal governments to encourage revision of exclusive and restrictive zoning laws prevalent throughout the suburbs? There is much that could be done to close the gaps that exist between and among the classes in America, but the powers that we are committed to the maintenance of the class system not the elimination of it.

It is now nine years since the passage of the Economic Opportunities Act and poverty has not been eliminated, society has not been restructured and with the advent of Nixon, the "total War on Poverty" has died an almost total death. Once survival needs are satisfied, poverty, after all continues to be a relative entity. Should the antipoverty programs have managed to raise the standard of living of the poor, as defined presently by society in terms of minimum income level, they would still have not achieved elimination of the inequities of income distribution in this country. However, if we have learned to regard the concessions of the system in terms of their real, not stated objectives (i.e., as a preventive measure against the coalescence of the masses that was feared might lead to a redistribution of income and power) and not naively as a panacea through which structural change is possible then we will have learned finally, to use the tools of the power structure and not be used by them. In this light one must then come to the paradoxical conclusion that the antipoverty programs are a success for both establishment and anti-establishment forces. In that the elimination of organized and disruptive "street" activity was effected (temporarily) by the antipoverty programs, the establishment temporarily succeeded. In that large numbers of radical personnel while ostensibly engaged in antipoverty business used this opportunity as a base around which to organize people and build political consciousness they too succeeded. For there are more people now than before that will never again be seduced into thinking possible the reformation of this system which distributes wealth without regard to need and manages humans without regard to humanity.

We must continue to build that base of organization; we must continue to foster the growth of political awareness until the time when the might of the establishment is not an impregnable barrier, but an obstacle that must and will be overcome.

Evaluation Comments

Jean Conroy was interviewed by several sociology faculty members. The following are quotes taken from their evaluation.

Faculty Evaluation A:

"Through her involvement with Open City as demonstrated in her essay and letter of recommendation from Nancy Fischer, Project Director of the Open Housing Program, Ms. Conroy has gained an understanding of the basics of Social Welfare.

As a Housing Counselor from 4/67 to 8/71 Ms. Conroy was involved in interviewing of clients, determining housing needs, discussing opportunities and alternative, counseling, and explaining procedures involved in discrimination cases and the clients' rights regarding housing. Through a personal interview I am satisfied that Ms. Conroy understands the history of social welfare in the United States and the social philosophy accompanying its growth."

Faculty Evaluation B:

"Her work experience more than exceeds the requisites for a field experience in social service. Her essay and a personal interview have demonstrated to me that she has competence in analyzing various social service agencies, their structures and modes of operation. Furthermore, she understands the dynamics of interviewing in psycho-social problem solving and is able to explore the common human needs through on-the-job experience."

Faculty Evaluation C:

"From 5/64 to 8/70 Ms. Conroy has led workshops on how to test and prepare a case of discrimination and workshops for volunteers of organizations on the role and responsibilities of white checker in cases of discrimination. She has an extensive list of community organizations she has been involved in since 1963 including: FOCUS (Civil Rights Organization); Big City CORE; The Freedom Democratic Movement and Party; Shorefront Committee for Nest Rocket Housing: Women Strike for Peace and a Pilot Project for Quality Integrated Education. Through a personal interview Ms. Conroy demonstrated her ability to plan, coordinate and implement direct action programs."

October 18, 1973

To Whom It May Concern:

Ms. Jean Conroy was employed by the Big City Urban League, Operation Open City from April 1967 through August 1971.

The purpose of this program, which has since become the Open Housing Program, is to assist minority families to Ms. Conroy, obtain the housing of their choice, according to the Fair Housing Laws of the City, State and Federal government.

Ms. Conroy's job title was Housing Counselor and encompassed the following duties:

Basic interviewing of clients, determining housing needs, discussing opportunities and alternatives, counseling on how and where to look, explaining procedures involved in discrimination cases and the clients' rights regarding housing. Where necessary, more in depth counseling was brought into play when there was evidence that clients needed information or referral on matters pertaining to education, employment, day care, health services, etc. This meant an extensive knowledge on Ms. Conroy's part of the sources and scope of these services. She acted as liaison between clients and the Department of Social Services, having acquired knowledge of the agency's procedures, structure, and policy relating to housing and budgetary problems. Ms. Conroy also had wide experience in handling housing discrimination cases--in their preparation and presentation to the City and State Divisions on Human Rights. In addition, Ms. Conroy was deeply involved in the planning and preparation of housing cases involving real estate brokers and salesmen which ultimately brought precedent-setting decisions from the Office of the Secretary of State.

Ms. Conroy recruited and instructed volunteers in the procedures of our program and supervised their activities during their involvement.

Ms. Conroy is knowledgeable in a variety of areas in the field of housing and social service, including public housing procedures, operational governmental programs and how to use them, and most importantly--how to bring all this knowledge in a useful and clear manner to clients from a variety of backgrounds who are seeking assistance and understanding.

OPEN HOUSING PROGRAM

Nancy Fischer

Nancy Fischer
Project Director

NF/ls

Faculty Portfolio Evaluation Report for Study Bl
Illustrative Portfolios

Portfolio Number __________

Faculty Assessor __________

Area of Expertise __________

Examine the entire portfolio carefully. Since the portfolio may contain evidence of prior learning in several academic or learning areas, you have been asked to award credits for that aspect of the portfolio and student learning which most nearly represents your area of expertise. For the purposes of this study there is no limit on the number of credits which can be awarded for the student's learning. Although it would be ideal to have the opportunity to interview the student or to obtain additional information, you are asked to evaluate the evidence as presented here.

1. This student has been awarded __________ credits by me in the area(s) of ____.

2. What specific material or information in this portfolio was most significant to you for your evaluation?

3. What material, if any, was NOT significant to you for your evaluation?

4. What specific criteria or benchmarks were used by you to determine the credit award?

APPENDIX 41

Institutional Portfolio Summary Sheet for Study B1
Actual Local Portfolios

	Credits	
Portfolio #	First Judgment	Second Judgment
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		

Fifteen Illustrative Examples of Prior Learning Outcomes For Study B2

1.

Learning outcome

Knows the roles and functions of the Midway Y's Men's Club and the Midway Civic and Commerce Association and the effect of these groups on the Midway Community.

Experience from which learning resulted

Completed independent study including participation as a member in the Midway Y's Men's Club and the Midway Civic and Commerce Association, reading its publications, personal interviews, and attending meetings and engaging in the groups' special functions.

Evidence of learning

Evaluation by the President of Midway Y's Men's Club, and the Executive Director of the Midway Civic and Commerce Association, based on an interview, indicated that the student shows a comprehensive grasp of the roles and functions of the Midway Y's Men's Club, its structure and how the club fits into the community and YMCA spectrum.

2.

Learning outcome

Knows and can apply the principles of cost accounting.

Experience from which learning resulted

Completed a course in cost accounting at a vocational school. Course covered practices, principles and procedures of handling production costs for use in inventory valuations and income determination, job order, process, and standard cost systems. Student was employed as a Cost Accountant at the Veterans Administration Hospital for nine years.

Evidence of learning

Received an outstanding performance rating in cost accounting from the Assistant Chief, Fiscal Division, Veterans Administration Hospital. Passed the Civil Service test qualifying her as a professional accountant. On the basis of past performance, was evaluated by her former supervisor now Chief, Fiscal Division, Veterans Administration Hospital. Summary evaluation:

a. is qualified to perform supervisory responsibility in the field of accounting, including cost accounting.

b. is qualified to assume the duties of Cost Accountant and Budget Technician relating to fiscal divisions in the Veterans Administration organization.

3.

Learning outcome

Knows and can apply techniques of written business communication.

Experience from which learning resulted

Over a period of 24 years in business management wrote frequently on subjects of sales promotion, sales proposals, manufacturing specifications, and other matters encountered in the manufacturing business. Also wrote internal memos, employee directives, operational instructions, analysis reports, and other documents common to business operations.

Evidence of learning

Written reports, letters and business documents were evaluated by a management consultant and a president of a manufacturing company. Summary evaluation:

a. Communications were of professional quality in construction and expression.

b. Effectively conveyed the intended messages.

4.

Learning outcome

Knows and can apply various principles in meal planning and serving.

Experience from which learning resulted

Collected menu file and new nutritionally sound recipes and menus. Planned and prepared budget meals designed to meet various developmental needs of family members. Entertained in formal and informal settings and taught cooking to family and friends.

Evidence of learning

Evaluation completed through submission of menu file, observation and interview by a faculty member at a community college indicated that the student showed basic understanding of principles and procedures involved for nutritionally and economically sound menus. The faculty member has instructed student nurses in basic and therapeutic nutrition for 11 years.

5.

Learning outcome

Knows and can apply techniques of fiction writing and can evaluate fiction written by others.

Experience from which learning resulted

Has written two novels, a book of poetry, short stories and is frequent guest book reviewer for the local paper.

Evidence of learning

One of the student's books is being considered for publication by a large national publishing firm. The book editor assigned to his manuscript reported that the student's book has a high probability of being accepted for publication in the upcoming year. His poetry has been published by a local publisher. Samples of the student's work were submitted to a creative writing professor. Summary evaluation:

a. His book reviews reveal keen skills in literacy criticism.

b. Novels and short stories show creative talent as well as skills in plot construction, description, and character development.

c. Written products show that the student is versatile in fiction writing in that samples cover the short story, novel, and poetry. Furthermore, they reveal that the student can handle suspense, humor, and satire in the written mode.

6.

Learning outcome

Knows non-technical German language skills and can apply them in communicating orally with native German speakers.

Experience from which learning resulted

Completed correspondence course in conversational German including the execution of written exercises and taped oral conversations which were assessed by correspondence school faculty. Works in a convalescent home which has a large number of German-speaking patients.

Evidence of learning

A college instructor's evaluation, based on exercises and oral conversations, indicated that, "she has shown a high degree of comprehension and a fine pronunciation of German. She would be comfortable speaking German with a native German in most social situations. She also has skills in translating written German. Written expression in German is somewhat weaker that her other skills."

7.

Learning outcome

Knows chief causes of abnormal personality development and therapeautic approaches commonly used in dealing with deviant behavior.

Experience from which learning resulted

Student reported reading three standard textbooks and six related books in the general subject area of deviant human behavior including: Abnormal Psychology and Modern Life (Coleman), Abnormal Psychology, Changing Conceptions of Psychoanalysis (Freud). Books were recommended by his psychologist and discussions as described by student were conducted between student and psychologist on content of books.

Evidence of learning

Requested and was given an oral examination on the topic of abnormal psychology by a college professor. This examination indicated that the student was able to define the principle causes of deviant behavior and compare several therapeautic techniques used in dealing with such behavior.

8.

Learning outcome

Knows and can apply concepts and skills of production typewriting, reproduction processes, and office supervision.

Experience from which learning resulted

Successfully completed three related courses required 30 contact hours each in production typewriting, reproduction, and office supervision at the Approved Business College. Worked for seven years as a secretary and office supervisor supervising three other employees.

Evidence of learning

Observed by three office supervisors in the process of typing, reproduction of materials, and supervision of office personnel. A standard rating form was completed and a judgment made by the three examiners. Summary of the rating form indicates the following:

a. Normal typing speed is 80 w.p.m.

b. Can efficiently use reproduction machines and make appropriate selection of reproduction process relative to assigned tasks.

c. Effectively organizes tasks, communicates expectations to subordinates, and objectively evaluates work of subordinates.

9.

Learning outcome

Knows and can apply basic principles of group dynamics in a work setting.

Experience from which learning resulted

Completed a company course in Interpersonal Communications III - Group Effectiveness, taught by her supervisor; participated in group discussions and demonstrations. Student reported reading assignments.

Evidence of learning

Based on observations, instructor reported that student demonstrated progression toward and attainment of a satisfactory level of competency in the following skill areas:

a. identification of the task and maintenance roles the members of a group have or have not assumed;

b. identification of helpful or hindering behaviors of a group's members in decision making;

c. identification of the leadership style a leader has assumed: telling, selling, testing, consulting, or joining;

d. identification of consequences of a group having clear versus unclear goals.

e. use of paraphrasing, perception checking, behavior description, description of feeling, and giving feedback as skills aimed at furthering communications and improving group process.

10.

Learning outcome

Knows the policies and procedures governing compliance with Equal Employment Opportunity (EEO) laws and can apply these to an affirmative action program in a major corporation.

Experience from which learning resulted

During a three-month internship served as Administrative Assistant to General Manager of the Corporate Equal Employment Opportunity and Affirmative Action Division, Econ Data Corporation. Administered various phases of the company's Affirmative Action program, interpreted company policies, and explained procedurally the intent of policy administration, which required understanding of the relationship of the company's EEO policies to its broader personnel policies. Assisted a division in another state in developing a company transportation program to solve an EEO problem. Investigated a formal charge of sex discrimination brought by the State Human Rights Commission. Assisted in a compliance review. Reviewed the work product of two professional administrators. Completed a budget review relating the department's budget to its functions in administering the corporate-wide EEO program.

Evidence of learning

On the basis of her performance during the internship was judged by her superior. Summary evaluation:

a. demonstrated knowledge of EEO laws and regulations and Econ Data's policies and procedures relative to EEO.

b. demonstrated ability to apply knowledge of EEO law in reviewing race and sex discrimination complaints and developing recommendations to resolve them.

c. demonstrated ability to apply knowledge of EEO law in solving management problems in situations involving employee relations.

d. performed outstandingly as an Administrative Assistant, demonstrating excellent administrative skills, keen analytical abilities and excellent judgmental skills.

e. because of performance during internship will be considered for permanent employment in a professional position.

11.

Learning outcome

Knows, can apply and evaluate the basic tenets of the major religious groups in the U.S.A.

Experience from which learning resulted

Over a five-year period, participated once a week in a discussion group sponsored by the Trinity Lutheran Church. This group read one book a month written by proponents of various religious groups in America and, under the direction of various leaders, discussed the meaning, implications, similarities and differences of various religious groups. Each participant was required to role-play the part of a practitioner of several religions and write an essay on her/his own religious beliefs with supporting rationale in relation to life goals.

Evidence of learning

Was rated by three discussion leaders as to knowledge of options in religious beliefs, ability and willingness to apply knowledge to his own life, and consistency and logic of arguments used to support those beliefs adopted by the student. A composite summary of the panel's evaluation was that the student did use informed judgment in selecting and/or formulating operational religious beliefs.

12.

Learning outcome

Knows and can apply skills of planning and conducting group activities for children, adolescents, and adults, including recreation, community involvement and therapy.

Experience from which learning resulted

Worked as a group worker volunteer with second-grade males at Bluff Multi-Service Center. Discussed group work programs available for children with the Center's Director, and consulted the principal and teachers at Stein Elementary School regarding activities that would benefit the boys. Met weekly with the boys for recreational and educational activities. Worked as a Supervisor at the Bear Lake Teen Center for one year. Planned and conducted programs and activities agreed on by teens and staff, counseled and referred teens with problems, and supervised adolescents in a teen drop-in center. Served as an intern with County Court Services and assisted the Volunteer Coordinator with the planning and conducting of orientation for adult volunteers. Has read texts on adolescent psychology, deviant behavior, human growth and development, and counseling.

✓

Evidence of learning

Observed and evaluated in planning and conducting group activities by Youth Program Coordinator, Bluff Multi-Service Center; Youth Counselor, Bear Lake Youth Resource Bureau; and Director, Volunteer Probation Officers' Program, County Court Services. Summary evaluation:

a. developed appropriate recreational and educational activities for children, and facilitated their independent pursuit of recreational and educational activity after his departure.

b. gained and demonstrated knowledge and understanding of the skills involved in planning, conducting, and supervising teen activities.

c. was an excellent supervisor and built positive relationships with teens, who respected his leadership.

d. his ability to communicate ideas gleaned from various resources was significant in the developing and implementation of programs.

e. was perceptive in assessing the relevance of various elements of volunteer orientation and training.

f. was able to apply concepts and theories gained from self-study and reading to practical counseling situations.

13.

Learning outcome

Knows the function and application of standard metal fabricating machines.

Experience from which learning resulted

Learned the functions of metal fabricating machines by observation, actual operation, and reading of instructional materials over a thirty-year period. Attended Wiedemann School for point to point programming of numerically controlled machines for 32 classroom hours.

Evidence of learning

While employed as a Division Manager of Manufacturing was observed by the company president. Summary evaluation:

a. Ability to determine manufacturing capabilities and estimate production time was dependent on knowledge of the capabilities of fabricating machines.

b. Knows the functions and applications of standard fabricating machines to an extent appropriate for a professional manufacturing manager.

14.

Learning outcome

Knows, can apply and can evaluate the basic theories methods of piano and accordion playing.

Experience from which learning resulted

Student reported attending weekly lessons in piano and accordion over a ten year period, reading a text on harmony, reading and studying various music technique books, practicing a minimum of one-half hour per day, teaching basic accordion lessons for one year, and teaching three children to play basic piano.

Evidence of learning

Letter from former employer certifies that student did know, apply and evaluate basic theories and methods of accordion playing while employed at Carleton Music Company. Was observed by two music professors at a local college, and they stated that student demonstrated basic theories and methods of accordion and piano playing and understood the application of these theories and methods.

15.

Learning outcome

Knows and can apply management and administrative skills in a church setting.

Experience from which learning resulted

Served as Assistant to the Pastor of the First Baptist Church. Organized and managed social functions involving college faculty and guest speakers from the United Sates and foreign countries. Reviewed new contacts for the church, determined process for orientation into the church and supervised follow-up on new members. Supervised the work of the custodian for a parish house. Supervised activities of young people and adults throughout a week-long committee meeting to assess needs. Assisted the Pastor in developing, organizing, and implementing congregational get-togethers relating to deeper levels of awareness and relationships. Reviewed data accumulated from meetings and disseminated information to staff for goal-setting purposes. Created and implemented a business/professional group for unmarried adults, including counseling members in such areas as housing, work-problems, spiritual life, and personal and social relationships.

Evidence of learning

Was observed by Director of Christian Education, First Baptist Church. Summary evaluation:

a. was competent in all areas of responsibility assumed.

b. was successful in organizing and implementing the various programs undertaken.

Faculty Recording Form for Study B2

Each of the 15 illustrations of student learning outcomes is accompanied by a description of the life experience from which it resulted and a brief evaluation statement. You are asked to determine the level of learning each of the outcomes represents by using the scale below. Please circle the number of the statement which indicates your judgment. In addition, indicate briefly reasons for your determination.

0 = Not worthy of college recognition.
1 = Worthy of lower division recognition (freshman-sophomore).
2 = Worthy of upper division recognition (junior-senior).
3 = Worthy of college recognition on the graduate level.

Illustration #	College Level	Reasons for Judgment
1	0 1 2 3	
2	0 1 2 3	
3	0 1 2 3	
4	0 1 2 3	
5	0 1 2 3	

Illustration #	College Level	Reasons for Judgment
6	0 1 2 3	
7	0 1 2 3	
8	0 1 2 3	
9	0 1 2 3	
10	0 1 2 3	
11	0 1 2 3	
12	0 1 2 3	
13	0 1 2 3	
14	0 1 2 3	
15	0 1 2 3	

Fifteen Illustrative Examples of Prior Learning Outcomes
For Study B3

1. I have been a musician from the age of 12 to the present. It has been a very rewarding profession, not only asthetically, but socially and financially as well. I own and play alto, tenor, and baritone saxophones, clarinet, flute and piccolo.

I have played many styles of music--classical, jazz, and rock. I have performed with chamber music quartets, small band groups, concerts and large band organizations. In addition I have taught saxophone, clarinet, and flute lessons, both privately and for studies.

I have read many types of music at sight. Music for shows, Classical music (Monmouth Symphony Orchestra, 1937-1940) (Monmouth Symphony, 1940 to 1943), Big Band Orchestrations (presently a member of Big Sound of the 70's, a 20 piece jazz band); Woodwind Quartette, (U.S. Army, on flute).

Perhaps this sounds repetitious, but I have had a great deal of experience in reading music, ear training, pitch and sight reading. Many times I have to work with musicians and bands in which we have never seen one another. This, as you know, tests musicianship to the fullest. I am called upon to perform with discretion in order to blend harmoniously with musicians around me in order to achieve the best musical effect possible.

2. In 1970 I attended a seminar titled "Principles of Supervision" sponsored by Cornell University. The seminar consisted of eight sessions.

All participants were asked to contribute, by listing their own major problematic areas of management. After the instructor presented diverse principles in management, we applied these same principles to the managerial problems. The knowledge of varied approaches to one problem was advantageous to me. The area that I obtained the most assistance in was evaluation of the staff personnel.

As a result of the course, I am more competent in evaluation interviews which I must perform in my capacity as Head Nurse of the Wycoff Hospital Emergency Room. I have a positive counseling attitude. The outcomes of the interviews are generally rewarding to the employee and myself. Constructive ideas concerning work in my department are carried out.

3. For many years I have been fascinated by the history of the American West with particular interest in the American Indian and the cowboy. This has been one of my hobbies. I have read historical texts and fictional works, viewed documentary films, and led library discussion groups on the topic of the American Indian and cowboys.

From these activities, I feel that I have gained a basic understanding of the Indian and cowboys in American history and culture. Furthermore, I have become aware of cowboy folklore and can distinguish between fact and fiction when I read fictional accounts and view films and television shows on the topic.

4. In my job as General Manager of the Manville Manufacturing Company I have served as a member of three labor contract negotiating teams and as chief negotiator for one of them. I was responsible for day-to-day management of union relations with three different unions for over a ten year period. I served as an advisor on union relations to corporate management for six years.

These experiences have given me extensive knowledge of negotiating labor union contracts, employee grievances, and other methods of maintaining rapport with labor bargaining units. Clearly, I have had the opportunity in my work to apply this knowledge in a practical setting.

5. I have successfully completed courses in Senior Lifesaving and Water Safety Instruction. I have read related textbooks, swim a minimum of one hour a week. Also, I have taught my daughters basic swimming techniques.

As a result of my coursework I feel I know and can apply the basic theories, principles, and techniques of swimming, life safety, and water safety. I am also aware of new materials and changes pending in these areas.

6. In 1968 after two brief attempts at a formal education at Glaston Monastery (1966 - 1967) and Rockbury Community College (1967 - 1968) certain economic and domestic deterrents arose which prevented me from furthering my formal education. Though an education in an institutional setting had become impossible, I did continue my studies of philosophy on a totally independent basis. In 1971 these studies resulted in my first book: BEYOND FIRST PRINCIPLES, Philosophical Library Inc. Such an accomplishment, however, produced a certain fear that perhaps I was now knowledgeable enough to defend such a work before learned critics. Perhaps it was this that caused me to accelerate my studies even more. During the course of the next two years I literally read every major work in western philosophy in addition to writing a second brief work entitled AN INQUIRY INTO PRAGMATIC IDEALISM.

I feel that I obtained an understanding of ethical philosophy as a result of my readings and writing endeavors.

7. I was cofounder, in 1946, of the Parker Research Corporation, served as Executive Vice-President until 1962, and then as President until 1968 at which time my partner and I sold the company. The company was mainly

engaged in opinion research as it relates to the sales effectiveness of advertising and audience appeal of programs on television and radio.

The Making of the President by Theodore White describes the results of a survey we conducted on the public's reaction to the Kennedy-Nixon debates. I developed the technique that was used, the questionnaire, and personally supervised the entire project. This gave me the opportunity to apply skills and creativity which I developed over the many years that I spent in the opinion research field.

8. Following graduation from State Technical College of Health Related Professions with an AAS, I received a faculty appointment as an Instructor in the School of Respiratory Therapy at this institution in February 1971. In this full time teaching position I learned of the need for planning and organization of teaching material. Effective teaching came slowly, primarily through practical development of skills on a trial and error basis. I learned that teaching is more than having answers and requires the application of supervisory and evaluative skills, particularly when engaging in clinical performance instruction at the bedside.

Also a member of the Executive Committee of the Inter-Agency Council on Smoking and Health, lecturer on smoking and chronic lung disease, and Instructor for the Heart Association. I have engaged in teaching a broad spectrum of students including area junior high and high school students, volunteer firemen and secondary school teachers. With such a varied group of learners, I found I could not use the same language and techniques of teaching that I was accustomed to at State Technical. I had to learn how to explain principles of science and physiology in both simplified and sophisticated ways.

I feel my teaching experience has led me to learnings in lecture organization, sequencing of material, testing and measurement of learning, development of course objectives and behavioral objectives, development and use of audio-visual techniques, and an increased understanding of the learner's needs.

9. I have been involved in the area of criminal justice since the summer of 1972, first as a Visitor and then as Coordinator with the Ecumenical Jail Visiting Group with functions under the aegis of the Metropolitan Church Board. I began readings in criminology in preparation for involvement with E.J.V.G.

Through reading, discussion with law enforcement officials, and a 10-hour training orientation, I have developed my understanding of the theory and philosophy of a Jail Ministry; legal implications of a Jail Ministry; role of counsel to incarcerated clients; organization and administration of criminal justice systems and agencies (police, prosecution, courts, corrections and rehabilitation); probation and parole systems (objectives, organization, practices, and assessment of effectiveness); bail system; pre-trial release programs; police-community relations and analysis of the problems involved in the formulation of ethical standards for police departments; causes and treatment of juvenile offenders by the police; social factors in crime and theories and methods of crime prevention; and the history of punishment.

10. In the nine years I spent as a clinical nurse at the State Veterans Home, under the close supervision of Dr. Robert Stone, I was able to build on my basic RN skills.

Using these skills I was able to assess the geriatric patient as a whole and to provide care as required. In being able to assess these individuals accurately, many facets were used, such as: theory, pharmacology, and physiology. This experience as a clinical nurse, in a geriatric setting, broadens my sound base for future work in geriatric administration, as well as that of being a well informed consultant.

This clinical nurse experience is comparable to the nurse clinician program in that I did learn assessment skills, theory of management and common health problems, pharmacology, pathology, physiology, as well as receiving much clinical practice. My clinical nurse experience was not the broad type which involves all individuals, but was very concentrated on the people I am interested in--the geriatric individual. Therefore, I feel more capable of caring for the geriatric patient than most nurse clinicians.

11. I am returning to college be become a better pastor and to meet the academic requirements to be received into associate membership of the Northern State Conference of the United Methodist Church. For this reason, I am requesting prior learning credits in the area of pastoral counseling.

In 1972 at the request of local clergymen, Dr. John Brent, Supervising Clinical Psychologist for the County Mental Health Clinics, organized a clergy consultation group. We met every three weeks to present case studies and discuss counseling situations. As a result of my participation in this group I developed some deep insights in interpersonal relationships. This group also sparked within me a deep interest in psychology. I spent approximately 50 to 75 hours overall in this group. A Methodist pastor is required to meet with couples for premarital counseling. I have met with approximately 50 couples over the past five-years and have spent approximately 100 to 150 hours in premarital counseling. In premarital counseling I use "The Marriage Role Expectation Inventory" published by Family Life Publications, Inc. The "Marriage Role Expectation Inventory" helps each partner to understand and examine their own marriage expectations in terms of reality and to recognize inconsistencies.

In premarital counseling the Inventory is of value as a part of departure in guiding couples to consider differing marriage expectations that may require reconciliation or understanding. The Inventory consists of 71 questions covering the areas of authority, homemaking, care of children, personal characteristics, social participation, education, and employment and support.

At the present time I am meeting with a group of clergymen every other Thursday. Each member of this group takes a turn at presenting a case study. I have spent about 15 hours with this group so far.

12. In December of 1969, I joined the Business Systems Division of Informatics Inc., a computer software company. The Business Systems Division, primarily, designed and implemented systems for clients in other companies. Because of the nature of this work, I was exposed to a variety of computer systems and software applications. The exposure improved my computer science expertise and my knowledge of various functions performed in a business environment. Some of the functional areas were personnel, accounting, materials management, engineering configuration management and customer relations.

The major software applications in which I participated are:

(1) Project leader and primary designer for an engineering configuration management system. This system, which operates in a batch environment on an IBM 370/165, is designed to maintain the latest configuration of end-item configuration.

(2) Design and implementation of a requisition work order and billing system. Operating in a batch environment, the system calculates and produces invoices for clients, based on the work type performed. In addition, the completion of the requisition is measured, which allows management to analyze performance.

(3) Implementation of a personnel system, which included substantial modifications to the original system design. The system maintains basic personnel data, performance reviews, a skills inventory, training (required and received), and a work history for each employee.

(4) Implementation of a parts control system, which faciliated ordering, re-ordering and client billing.

(5) Provided training and implementation guidelines for a company that had recently acquired MARKIV.

(6) Design and implementation of a customer inquiry system. This system initiated the investigation, reported the on-going status and provided management performance reports.

Other systems in which I participated include manpower planning, project control, salary surveys and accounts receivable.

13. As the owner of a small manufacturing plan, I decided to undertake a study of the Cooperative Wage Study plan for job description and classification developed by the U.S. Steel Corporation. I visited a local plant which utilizes this plan and observed its efficacy. It was then that I decided to develop a program along these lines for my own plant which has recently gone into operation.

14. My life experience dealing with criminology has been concentrated in the field of law and public safety. I have been a part time police officer in the communities of Butler and Riverdale since 1955, and have performed this service for a minimum of eight hours per week for twelve years. During this time, I have spent many hours in lecture sessions, presented by the State Police, Civilian Defense and Disaster Control, Morrow County Prosecutor's Office, and the Federal Bureau of Investigation. In addition, several lectures were presented by the local magistrate so that each officer would have more knowledge and a better understanding of court justice, case preparation and presentation, and penal institutions.

During my period of service, I have been exposed to police patrol for motor vehicle violators where knowledge of State Motor Vehicle and Traffic Regulations, Title 39 is mandatory. In addition, a working knowledge of state statutes and municipal codes is also a must. Even though major crimes are not often committed in the local community, an officer has to be prepared to lodge the proper criminal charge against a perpetrator. By the same token, an improper charge or a charge against the wrong individual can result in a libel lawsuit against the officer and municipality.

Police training has taught me to understand community and social problems typically represented by marital problems, family disputes, disorderly conduct of persons, crowd control, drug abuse, juvenile delinquency and intoxication.

Crime is often propagated by an individual who has been alienated by our society. Training and practical experience has given me a better understanding of what makes a criminal and how to work with individuals who are belligerent towards our society. Understanding them and being able to help them to help themselves will make our society a better place to live.

My appearances in municipal and county courts has given me a good understanding of American justice. Some of my police experience has gone into aiding local prosecutors for court hearings while other times I have just been an interested observer in court.

15. I participated in a 10-week workshop (2-hour sessions) in Parent Effectiveness Training offered through the Adult Education Division of West Genesee School District. The text used was Dr. Thomas Gordon's book Parent Effectiveness Training: The No-Lose Method of Raising Responsible Children. Topics studied were barriers to communication, inhibiting and facilitation responses, active listening, feedback, levels and language of acceptance, theory of communication, the child-parent power struggle, defining ownership of a problem, win-lose and no-lose approaches to problem-solving, values conflicts, and conflict resolution. Dr. Gordon's methods were cited as a promising new model for the prevention of family disintegration in the report of the 1970 White House Conference on Children. My supplemental readings in child and adolescent psychology added to my learning of the theory of communication, and I was able to apply this knowledge through my involvement in Room 112.

In the Spring of 1973, responsive to student unrest at the local high school, another mother and I met with some students and school officials to try to develop a program to meet needs of students which were not being met by traditional methods. Over the summer, through meetings with students, teachers, parents, and school personnel, the concept of Room 112 evolved. The school agreed to set aside a classroom to be used as an informal room where students could go to relax, study, or share a personal concern with a student trained in peer counseling techniques or a competent and interested adult from the community. The climate was unstructured, informal, accepting, and conducive to personal growth.

The room was staffed by students, teachers, community adults, and the school psychologist, Mr. Robert Burdick, for one day a week. I consider my P.E.T./Room 112 learnings and practice related to college-level courses in theory of communication and parent-child relations.

Checklist for Recording Type of Documentation Required for Study B3

After reading the 15 excerpts taken from student portfolios, determine the articles of documentation you would require of the student to assist you in assessing the type and level of learning that may have resulted from the life experience described. A list of various kinds of documentation is provided. Place a check in the box that best describes the documentation you would require for each example. You may require more than one type of documentation or, in some cases, none.

	Illustration #														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Letters of verification that experience took place from supervisor, client, or coworkers															
2. Letters of evaluation of performance from supervisor, client, or coworkers															
3. Letters of commendation															
4. Official job description															
5. Student explanation of tasks performed															
6. Licenses															
7. Certificates															
8. Membership in professional organizations															
9. Course outlines, syllabi, and/or assignments															
10. List of books read															
11. Products or work samples such as plans, tapes of performance, manuscripts, memos, reports, etc.															
12. Other*															

---

*Please describe your additional documentation requirements on a separate sheet of paper for the illustrations checked in this category.

Form for Recording Verification of Documentation for Study B3

At the top of each box write the number of documents of each type found in each Portfolio. In the bottom half of each box write the number of items which were determined by you to be authentic.

Portfolio #

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1. Letters of Verification																									
2. Letters of Evaluation																									
3. Letters of Commendation																									
4. Official Job Description																									
5. Certificates																									

Guidelines for Determining Document Authenticity for Study B3

Select 25 portfolios from your files and number them 1 through 25. Contact the students who are the subjects of these portfolios and explain to them that your institution is participating in a large national study which seeks to improve procedures in portfolio or prior learning assessment, and that an investigation of the importance and authenticity of documentation is included in the study. Ask each student if he or she has any objection to having you check the sources of documentation found in his/her portfolio (you may want each student to file a permission slip).

After obtaining student permission, examine each portfolio carefully and extract articles of documentation that correspond most closely to the categories listed below.

1. Letters of verification - letters, usually written by a supervisor, client, or colleague, which state only that a particular experience took place.

2. Letters of evaluation - letters, usually from supervisors or outside experts, which actually list and/or evaluate the learning outcomes.

3. Letters of commendation - letters, usually written by agencies, corporate heads, organizations, or clients, which commend a student for outstanding performance of a particular activity and are much like a letter of recommendation.

4. Official job description - a description which appears to come from a company's or agency's personnel division or department and not written by the student.

5. Certificates - documents which indicate the completion of an activity or satisfactory performance of a set of tasks.

The authenticity of the documentation found in the portfolios can be checked by a telephone call or correspondence to the source of the document. The choice of contact made depends on your budgetary constraints as well as your convenience.

Student Questionnaire for Study B4

Please answer each question as fully but as concisely as possible. Try to confine your answer to the space provided. If you need more space, attach additional sheets of paper with the question numbers indicated.

1. If you have had any previous experiences in work, community or church activities, hobbies, and/or travel which you feel resulted in learning that is worthy of college credit, please list them and what you learned from them below.

Experience	Learning Outcomes

2. For each of the experiences listed above explain how you could demonstrate to a faculty member of this college that you indeed learned what you said you learned.

3. Briefly explain how your life experience learning is related to your career goals.

4. Briefly explain how your life experience learning is related to your educational goals.

5. Briefly explain how your life experience learning is related to your personal goals.

4-81

Site Coordinator Questionnaire for Study B4

Check any of the choices below which apply to your institution or program.

1. Which of the following arrangements are provided by your institution to assist students in presenting prior learning for credit?
   a. counseling - group _______ individual _______
   b. mentoring or advising by faculty members _______
   c. student workshop (non-credit) _______
   d. educational planning course (credit) _______
   e. self-instructional materials _______
   f. other (explain) _______

2. Which of the following is the student required to produce in order to have prior learning assessed?
   a. a resume or vita _______________
   b. an autobiography _______________
   c. degree pacts or contracts _______________
   d. learning contracts _______________
   e. documentation _______________
   f. narrative _______________
   g. other (explain) _______________

3. Which of the following measurement techniques are used by assessors to measure the presence and/or extent of prior learning?
   a. product assessment _______
   b. interviews _______
   c. formal oral examinations _______

d. simulations _______

e. performance testing or observations of performance _______

f. essay tests _______

g. objective tests other than CLEP _______

h. self-assessment instruments such as student checklists or questionnaires _______

i. other (explain) _______

4. Who does the assessment of prior learning?

a. individual faculty member for each specific learning or discipline area indicated in the student's folder or portfolio _______

b. interdisciplinary faculty committee _______

c. outside experts _______

d. alumni experts _______

e. peers (students) _______

f. assessment team which may contain faculty, administrators, outside experts, alumni, etc. _______

g. other (explain) _______

5. How is learning evaluated and transcribed?

a. courses offered in the college _______

b. subject or discipline area _______

c. competencies _______

d. life learning credits - unspecified _______

e. other (explain) _______

Please attach a brief description of the process by which a typical student would obtain credit for or recognition of prior learning. Attach any institutional materials used by students and faculty which would supplement your description.

Revised Guidelines for Site Coordinators

Area C - Work Experience

The purpose of these guidelines is to provide research site coordinators with as much detail as possible concerning the background of the validation studies and the steps which should be taken to ensure consistency in experimental procedures and data collection. First, a brief description of the studies is given, then a series of steps is presented for each study which should be approximately followed at each field research site. Finally, an overall schedule for the validation phase is given.

Description of the Studies

Study #C1 - Do faculty agree on the amount of credit to be awarded for work experience?

The basic purpose of this study is to see to what extent independent assessors can agree on the amount of credit to be awarded for work experience. Using a local adaptation of the methods described in CAEL Working Paper No. 8, two different faculty assessors will independently estimate the amount of credit to be awarded for a minimum sample of 20 students. Data will then be analyzed so that conclusions can be drawn regarding the extent of agreement among faculty. CAEL Working Paper No. 8 dealt with three areas of work experience -- Data Processing, Law Enforcement, and Secretarial Science. Rating forms for each of those areas have been provided. In addition, a fourth area, Management, which was not treated in Working Paper No. 8, has been added. Each rating scale asks for estimates of credits within specific categories within each area.

The 20 students selected for the study should be as representative as possible of the normal range of students. Since not all institutions have programs in all four of the areas chosen the 20 students selected for study need not represent all four areas. All 20 could come from one area only or could be distributed across two, three, or four areas.

Study #C2 - To what extent can faculty articulate work experience within a particular occupational program?

The purpose of this study is to examine the extent to which faculty within an institution can agree on the relevance of specific competencies to areas within a particular field or occupational program. A total of five different portfolios in each of the areas of management, law enforcement, and data processing will be judged by at least three assessors with expertise in each area. Thus, a minimum of nine judges, three for each area, will be required. Each judge will be asked to read through the five portfolios in his or her area which will present evidence of prior learning through work experience and assign credit totals for specified subareas as well as a total amount of credit.

Study #C3 - What is the concurrent validity of the documentation of work competence?

The basic purpose of this study is to examine the extent to which students and their work supervisors agree on which competencies were actually learned on a particular job or in a particular work experience. A minimum of 20 students seeking credit for work experience must be included in the study. One of the three task inventories provided in Working Paper No. 8 will be filled out by each student and by the student's work supervisor. Data will be analyzed to determine the extent of agreement between student and work supervisor.

Study #C4 - What is the validity of advanced placement in occupational programs on the basis of work experience?

Students are often placed in advanced course sequences on the basis of prior work experience. The purpose of this study is to determine whether students so exempted are as generally competent as those who have gone through the relevant course. The performance and self-perceptions of exempted students will be prepared with the performance and perceptions of traditionally-placed students. Participants for the study should be identified from students in one or more advanced courses related to an occupational program. A minimum of 10 traditionally-placed students and 10 students placed on the basis of work experience will be required.

Steps in the Collection of Data

1. Check Materials Needed

All materials necessary for completing data collection have been provided. As a first step, however, the site coordinators should become familiar with the materials and make sure that an adequate supply of each item is available. Materials which will be supplied are as follows:

Study C1

- Fifty (50) faculty forms for judging credits to be awarded to actual students

Study C2

- Five (5) sets of 15 simulated cases
- Five (5) sets of faculty rating forms and instructions for each area

Study C3

- Twenty-five (25) copies of task inventories in each of three fields with instructions for students
- Twenty-five (25) copies of task inventories in each of three fields with instructions for work supervisors

Study C4

- Thirty (30) end-of-course questionnaires

2. Plan Data Collection

a. Obtain faculty and work supervisor cooperation: All four studies will require the cooperation of faculty or (in the case of Study C3) work supervisors. For Studies C1 and C2 a minimum of five faculty will be needed to judge the 15 simulated cases. In order to simplify data collection it would be desirable if one group of faculty could serve as judges for both Studies C1 and C2. For Study C3 work supervisors must agree to complete task inventories for students. Study C4 will require course instructors to administer a student questionnaire and provide a roster of grades for all students. The agreement of participants should be obtained as early as possible.

b. Obtain student cooperation: For Studies C1, C3, and C4 the agreement of students to use their records for study purposes must be obtained. Students selected should be as representative as possible of the normal range of assessees.

3. Meet with Faculty Participants

A meeting should be held with faculty participating in Studies C1 and C2 before data collection begins. The purpose of the meeting will be to acquaint faculty with the research studies, to answer any questions that are raised, and to give them specific details on how data are to be recorded and collected. One of the points which should be emphasized at this meeting is that the ratings or judgments must be made independently. That is, there should be no collaboration or discussion of the ratings made by different faculty members until after all ratings have been completed and returned to the site coordinator. Participants should also be informed at this time as to arrangements for collecting their completed ratings.

Since it is probably not feasible to meet with work supervisors to explain Study C3, the instructions accompanying the task lists will have to serve this purpose. Work supervisors should be encouraged to contact the site coordinator if they have any questions or problems, however.

A meeting is probably not necessary for Study C4, but each course instructor involved should be contacted individually to clarify procedures.

4. Schedule Data Collection

A schedule of data collection should be drawn up and distributed to faculty and work supervisors. Since the studies involving actual students may be difficult to schedule precisely only a rough projection of the time frame within which data collection will occur will be necessary. A copy of the data collection plan should be sent to the ETS coordinator.

5. Collection of Data

Before the 1st of January at least 1/3 of the data in each of the three studies should be collected. The remaining 2/3 of the data can be collected after January 1, but no later than the beginning of March.

6. Attend Meeting for Site Coordinators in Early January

A meeting will be held in early January for all site coordinators at which time any problems or difficulties which have been encountered in the collection of data can be raised and hopefully ironed out.

7. Return Data to Educational Testing Service

Data should be returned no later than March 1, 1976.

8. Schedule

The following are key dates for site coordinators in planning and executing the data collection.

Date	Event
Mid October	CAEL Assembly Meeting
Early November	Complete Plans for Data Collection
Mid November	Complete Meetings with Faculty Participants and Schedule
January 1	Complete at least 1/3 of the Data Collection in each of the Three Studies
January 8	Attend Meeting for Site Coordinators
March 1	Finish Data Collection and Mail to ETS

Work Experience Evaluation Form for Study C1: Data Processing

Faculty Assessor ______________________________

Student Name __________________________________

Based on all the evidence you have examined for this student, you are asked to indicate the amount of credit which you would award in each of the indicated categories (assume there is no ceiling on the amount of credit which can be awarded). Please do not discuss your assessments with anyone else. The data you provide will be used for research purposes only and no individuals or institutions will be identified.

Area of Work Experience	Total Credits
a. Operating Data Processing Equipment	________
b. General Computer Programming	________
c. Commercial Computer Programming	________
d. Scientific Computer Programming	________
e. Systems Analysis and Design	________
f. Data Processing Management	________
g. Other Data Processing Credits	________
h. Total Data Processing Credits	________

Work Experience Evaluation Form for Study C1: Law Enforcement

Faculty Assessor ______________________________

Student Name __________________________________

Based on all the evidence you have examined for this student, you are asked to indicate the amount of credit which you would award in each of the indicated categories (assume there is no ceiling on the amount of credit which can be awarded). Please do not discuss your assessments with anyone else. The data you provide will be used for research purposes only and no individuals or institutions will be identified.

Area of Work Experience	Total Credits
a. Police Operations	________
b. Criminal Investigations	________
c. Criminal and Constitutional Law	________
d. Criminal Evidence	________
e. Criminal Procedures	________
f. Administration	________
g. Other Law Enforcement Credits	________
h. Total Law Enforcement Credits	________

Work Experience Evaluation Form for Study C1: Secretarial Science

Faculty Assessor ______________________________

Student Name ___________________________________

Based on all the evidence you have examined for this student, you are asked to indicate the amount of credit which you would award in each of the indicated categories (assume there is no ceiling on the amount of credit which can be awarded). Please do not discuss your assessments with anyone else. The data you provide will be used for research purposes only and no individuals or institutions will be identified.

Area of Work Experience	Total Credits
a. Typewriting	________
b. Office Procedures	________
c. Record Keeping	________
d. Office Machines	________
e. Communications	________
f. Supervision	________
g. Other Secretarial Science Credits	________
h. Total Secretarial Science Credits	________

Work Experience Evaluation Form for Study C1: Management

Faculty Assessor ______________________________

Student Name ________________________________

Based on all the evidence you have examined for this student, you are asked to indicate the amount of credit which you would award in each of the indicated categories (assume there is no ceiling on the amount of credit which can be awarded). Please do not discuss your assessment with anyone else. The data you provide will be used for research purposes only and no individuals or institutions will be identified.

Assessors will notice the functions listed below as being the six basic management functions. This is not meant to be all inclusive and, of course, could take into account competencies evidenced by work experience subsumed under these functions such as problem solving, social responsibility of business, business law, information systems, taxes, accounting, marketing, economics, international business, personnel and labor relations, political context of business, decision-making, innovation, etc. The evaluator, of course, should be aware of the fact that in practice, these functions are iterative, i.e. each contains elements of the others, however, in a conceptual framework, they can be separated as suggested below.

Area of Work Experience	Total Credits
a. Planning	________
b. Controlling	________
c. Communicating	________
d. Organizing	________
e. Directing	________
f. Motivating	________
g. Other Management Credits	________
h. Total Management Credits	________

CASE C2 -- DATA PROCESSING II

Drysdale County Community College

Life/Learning Experience Assessment Program

Chronological Record

Date: May 6, 1975

Name: William Peterson

Please give a chronological summary of your life from high school to present. You should include work, study, community activities, and organizations, short courses and training programs, marriage, birth of children. In short, important events from which you have learned as you have come to where you are today. This is a "dates of events" list.

June-1954	Graduated from Springdale High School
June-1954 to Nov.-1960	U. S. Army
Dec.-1960 to Dec.-1964	Worked for American Baptist Publications
Mar.-1965 to Present	Worked for Snyder, Inc.

Drysdale County Community College

Life/Learning Experiential Assessment Proficiency

Autobiographical Document

Date May 6, 1975 Name William Peterson

Write an account which parallels the chronological record which you have prepared. The purpose of this document is to comment on and put into perspective the events of your life to date. As a part of this document, a section at the end should lay out future plans and how your studies at DCCC mesh with these plans.

After graduating from Springdale High School in June 1954, I entered the Army as an enlisted reservist. My last year of high school I had been in the active Army reserve and decided to get my military service over with. While serving my enlistment in various countries such as Korea, Alaska, and Japan I was an Analytical IBM keypunch operator in Army Security Agency and subsequently became a machine operator and a Morse Code Operator.

Upon my discharge from active service in November 1960, I obtained a job as an IBM 407 operator with American Baptist Publication Society. During the four years I worked for this company I learned to wire various machines and learned programming on a Burregouhs 101 computer. I was considered a lead or senior operator at the time of my leaving the company.

After leaving my position with American Baptist in December 1964, I obtained a job as an IBM machine operator with Snyder, Inc. in March 1965. During the first 3 1/2 years with Snyder, I operated an IBM 1401 Computer with two disk drives. Also I was named a night supervisor with other operators under me and also the night shift of the keypunch department.

In April of 1968 I took a test with 10 other employees for a chance of obtaining a job as a programmer trainee. I came out fourth highest and was selected for the programming spot. In September, after having trained my replacement, I entered the programming department of Synder. Since this time, I have learned 1401, 360 RPG, 360 Assembler Language, and DOS job control with a little SPS thrown in.

During the time from September 1968 to the present date I have risen from Programmer Trainee to my present position of Senior Programmer. Most of my programming training was obtained through the use of P.I. courses (Personal Instruction) which consists of books and work courses done by the student and then tests and workshops at the IBM Center. On-the-job training at work consisted of actually writing and debugging programs to be activated into the Snyder system.

Drysdale County Community College

Life/Learning Experience Assessment Program

Skill Inventory

Date May 6, 1975 Name William Peterson

In your work, training and experience, what skills and abilities have you gained? Please list these and indicate situations in which you have employed them. A few lines, or a short paragraph, on each of these skills and abilities will help the faculty evaluator assess your entry level into your studies. Please be extensive but as precise as you can.

During the time I was serving in the U. S. Army I obtained skills in wiring and running the following equipment:

IBM--082, 089, 101, 602A, 604, 514, 519, 402, 407, 548, 026. I went to various training schools conducted by the Army and IBM to learn these skills and had on-the-job training also. As a result of first having gone to six months of Morse Code Interceptor School the IBM training soon followed in conjunction with the Morse Code training and my subsequent assignment to Army Security Agency.

When I went to work for American Baptist Publications, I operated and wired the following equipment:

IBM--082, 514, 519, 602A, 407, and Burr. 101 Computer. American Baptist Publications sent me to school for wiring the 407 and I had a work-school type training for the computer programming for the 101. The work done for American Baptist was inventory, billing, accounting, and order filling through the IBM system.

During the time I was working for American Baptist Publications, I attended night school at Makefield University where I majored in Accounting. During two years of night school I obtained 12 credits in accounting and 3 credits in Business Law. The accounting came in handy for my job at work.

When I went to work for Snyder, Inc., I operated the following machines (and I still do in the course of my job as programmer):

IBM--1401 Computer, 360/30 Computer with three 2311 Disk Drives and two Telex 1600 BPI Tape Drives, 514, 089, 083, and Inforex Key to Disk data entry equipment.

My programming training has been obtained through PI courses and on-the-job training. The schooling at Makefield has helped me in my present position.

I am hoping that my present skill and training can subsequently lead to systems and design work and schooling at DCCC will advance me toward that goal.

March 6, 1975

Work History of William Peterson

To Whom It May Concern:

Mr. Peterson joined Snyder, Inc. on March 22, 1965, in the capacity of Second Shift Machine Operator. This position required him to operate all of the following pieces of equipment with large volumes of cards.

1. 083 Sorter
2. 514 Reproducing Punch
3. 548 Interpreter
4. 085 Collator
5. 1401 Computer with 1311 Disk

In July 1966, Mr. Peterson was promoted to the position of Second Shift Operating Supervisor. This position required him to maintain most of his previous duties, plus assist and supervise younger operators with less experience. In June 1967, Snyder installed a 360-30 with 2415 Tape Drives and 2311 Disk Storage. Mr. Peterson was required to learn the operating of these new pieces of hardware which he did very well.

In September 1968, Mr. Peterson was promoted to the position of Programmer Trainee. In this position, he was required to learn 1400 symbolic programming, 1400 autocoder, 360 assembler language, 360 RPG, and DOS job control. Upon completion of his education, he was required to maintain all existing programs.

In September 1969, Mr. Peterson was promoted to the position of Junior Programmer. In this position he retained all previous responsibilities, plus he was required to provide operator assistance when needed and became involved with new application work.

In July 1974, Mr. Peterson was promoted to the position of Senior Programmer. To qualify for this position, an individual must master all phases of programming. This includes flowcharting, coding, assembly, testing, desk debugging, and above all other, thorough documentation.

I feel this represents an accurate account of Mr. Peterson's growth path from March 1965 to the present.

Sincerely,

R. A. Sanford

R. A. Sanford
Director, Data Processing

RAS/ls

cc: file

Department of the Army

Certificate of Training

This is to certify that

PVT Peterson , ER 13481194

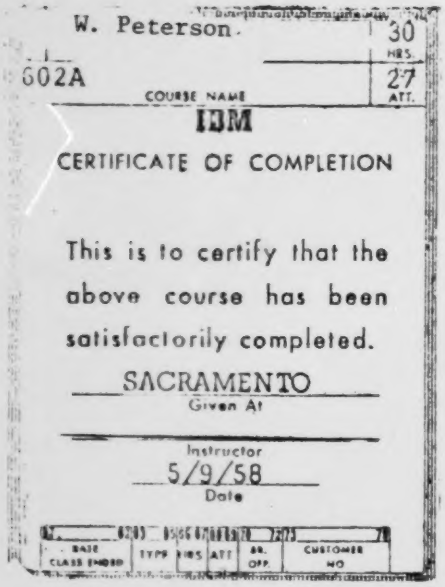
has successfully completed

THE 39 HOUR ANALYTIC EQUIPMENT CARD PUNCH OPERATOR COURSE
(MOS 41808)

Given at ARMY SECURITY AGENCY SCHOOL

26 MARCH 1955
(DATE)

RICHARD E. JONES
Lt. Colonel, Artillery
Assistant Commandant



W. Peterson — 30 HRS.

602A — 27 ATT.

COURSE NAME

IBM

CERTIFICATE OF COMPLETION

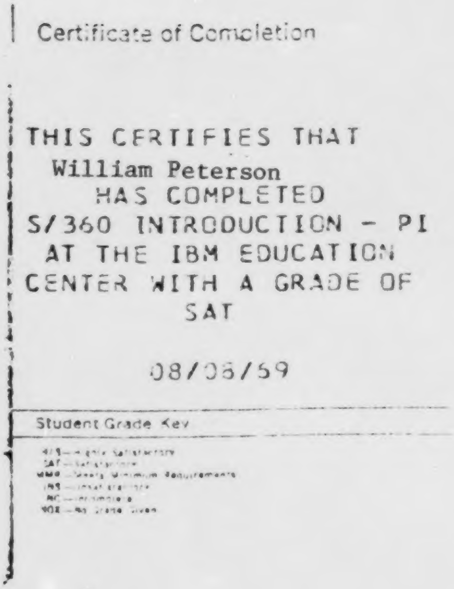
This is to certify that the above course has been satisfactorily completed.

SACRAMENTO
Given At

Instructor

5/9/58
Date

DATE CLASS ENDED | TYPE | HRS | ATT | BR. OFF. | CUSTOMER NO

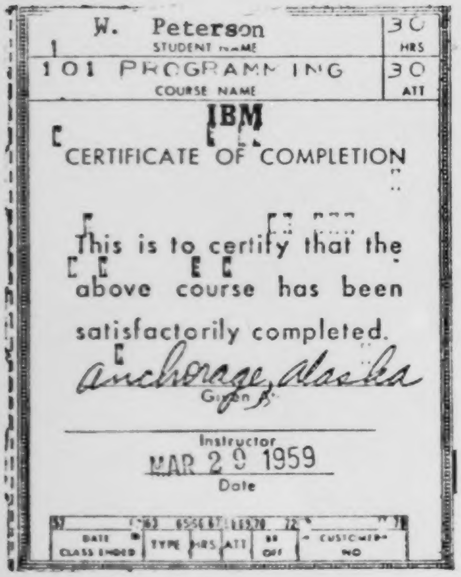


Certificate of Completion

THIS CERTIFIES THAT
William Peterson
HAS COMPLETED
S/360 INTRODUCTION - PI
AT THE IBM EDUCATION
CENTER WITH A GRADE OF
SAT

08/06/59

Student Grade Key



W. Peterson — 30 HRS
STUDENT NAME

101 PROGRAMMING — 30 ATT
COURSE NAME

IBM

CERTIFICATE OF COMPLETION

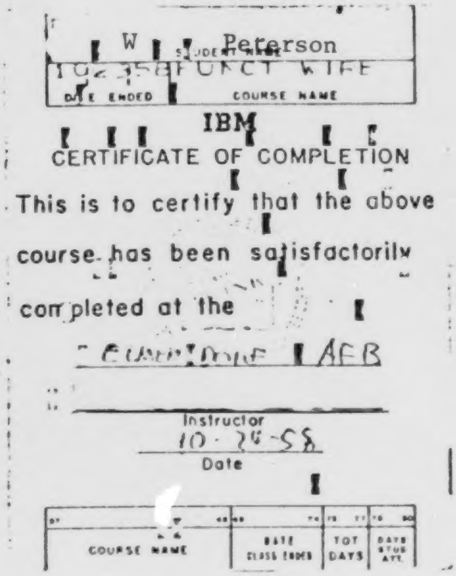
This is to certify that the above course has been satisfactorily completed.

Anchorage, Alaska
Given At

Instructor

MAR 29 1959
Date

DATE CLASS ENDED | TYPE | HRS | ATT | BR OFF | CUSTOMER NO



W Peterson
STUDENT NAME

[illegible]
DATE ENDED | COURSE NAME

IBM

CERTIFICATE OF COMPLETION

This is to certify that the above course has been satisfactorily completed at the

[illegible] AFB

Instructor

10-24-58
Date

COURSE NAME | DATE CLASS ENDED | TOT DAYS | DAYS STUD ATT

RECEIVED FEB 2 6 1975

MAKEFIELD UNIVERSITY

ADMITTED FROM Springdale High School
FIRST REGISTRATION 9/62

ENTRANCE UNITS	ENG.	HIST.	LAT.	FR.	GER.	SP.	MAT.	GEOM.	SOL.	TRIG.	S. SCI.	BIO.	PHY.	CHEM.	ETC.
	3	4					3				1	1			

Peterson, William John 63398
LAST, FIRST, MIDDLE NAME, NUMBER

PARENT-GUARDIAN
BIRTH PLACE Lansing, Penn.
BIRTH DATE JUNE 22, 1935

DEPT.	COURSE NUMBER	COURSE NAME	CR. HRS.	GR.	Q. PTS.
63398- ACCT	101	FALL 1962 PRINC OF ACCOUNTING I	3	B	9
63398 ACCT	102	SPRING 1963 N M PRINC OF ACCOUNTNG II	3	B	9
63398 ACCT	121	FALL 1963 N M PRINC ACCT INTERMED I	3	C	6
63398 ACCT	122	SPRING 1964 N M INTERMEDIATE ACCT II	3	B	9
63398 B L	109	FALL 1964 N M U S JURISPRUDENCE	3	B	9

DEPT.	COURSE NUMBER	COURSE NAME	CR HRS	GR.	Q PTS

This record is official only when duly signed and sealed.

Date FEB 25 1975

Joan M. Snow
(Mrs.) Joan M. Snow
Registrar

4-99

Faculty Rating Form for Study C2: Data Processing

Below are listed six ares in Data Processing within which credit might be awarded based on previous work experience. An additional category is provided for credits which might be awarded for experience outside those categories listed. You are asked to read each of the five cases carefully and estimate what in your best judgment would be the appropriate number of credits to be awarded in each of those areas and the total number of credits. You are asked to assume that there is no ceiling on the number of credits.

Case #	Operating Data Processing Equipment	General Computer Programming	Commercial Computer Programming	Scientific Computer Programming	Systems Analysis Design	Data Processing Management	Other Data Processing Credits	Total Data Processing Credits
D1								
D2								
D3								
D4								
D5								

Faculty Rating Form for Study C2: Law Enforcement

Below are listed six areas in Law Enforcement within which credit might be awarded based on previous work experience. An additional category is provided for credits which might be awarded for experience outside those categories listed. You are asked to read each of the five cases carefully and estimate what in your best judgment would be the appropriate number of credits to be awarded in each of those areas and the total number of credits. You are asked to assume that there is no ceiling on the number of credits.

Case #	Police Operations	Criminal Investigations	Criminal and Constitutional Law	Criminal Evidence	Criminal Procedures	Administration	Other Law Enforcement Credits	Total Law Enforcement Credits
L1								
L2								
L3								
L4								
L5								

Faculty Rating Form for Study C2: Management

Below are listed six areas in Management within which credit might be awarded based on previous work experience. An additional category is provided for credits which might be awarded for experience outside those categories listed. You are asked to read each of the five cases carefully and estimate what in your best judgment would be the appropriate number of credits to be awarded in each of those areas and the total number of credits. You are asked to assume that there is no ceiling on the number of credits.

Assessors will notice the functions listed below as being the six basic management functions. This is not meant to be all inclusive and, of course, could take into account competencies evidenced by work experience subsumed under these functions such as problem solving, social responsibility of business, business law, information systems, taxes, accounting, marketing, economics, international business, personnel and labor relations, political context of business, decision-making, innovation, etc. The evaluator, of course, should be aware of the fact that in practice, these functions are iterative, i.e. each contains elements of the others, however, in a conceptual framework, they can be separated as suggested below.

Case #	Planning	Controlling	Communicating	Organizing	Directing	Motivating	Other Management Credits	Total Management Credits
M1								
M2								
M3								
M4								
M5								

Faculty Instructions for Study C3

IV. Job Inventory Checklist

Listed on the following pages are specific tasks which are typically conducted by policemen. The task descriptions are divided into six major categories: A. Police Operations, B. Criminal Investigation, C. Criminal and Constitutional Law, D. Criminal Evidence, E. Criminal Procedures, and F. Administration. Read each category and indicate with a check (✓) which tasks were performed while under your supervision. Relative time means the total time spent doing the task compared with the time spent on each of the other tasks. Then enter the relative time spent on the task by using the code numbers given below.

CHECK (✓) ALL OF THE TASKS YOU HAVE PERFORMED IN THE ENTIRE COLUMN OF STEP 1 BEFORE YOU BEGIN TO ENTER THE TIME CODE FOR EACH IN STEP 2. ENTER TIME CODE ONLY FOR TASKS YOU HAVE PERFORMED.

TIME CODE

1. VERY MUCH BELOW AVERAGE TIME ON THIS TASK.
2. BELOW AVERAGE TIME.
3. AVERAGE TIME.
4. ABOVE AVERAGE TIME.
5. VERY MUCH ABOVE AVERAGE TIME ON THIS TASK.

(Time does not necessarily equal importance)

Example:	STEP 1 CHECK PREVIOUS JOB TASKS	STEP 2 ENTER TIME CODE
1. Issue moving traffic citations	✓	5
2. Issue warning tickets		
3. Administer field tests for intoxication (coordination tests, etc.)	✓	4
4. Check autos against stolen car list		
5. Recover stolen automobile	✓	2

Student Checklist for Study C3: Law Enforcement

IV. Job Inventory Checklist

Listed on the following pages are specific tasks which are typically conducted by policemen. The task descriptions are divided into six major categories: A. Police Operations, B. Criminal Investigation, C. Criminal and Constitutional Law, D. Criminal Evidence, E. Criminal Procedures, and F. Administration. Read each category and indicate with a check (✓) which tasks you performed in your previous work. Relative time means the total time you spent doing the task compared with the time you spent on each of the other tasks on your previous jobs. Then enter the relative time you spent on the task by using the code numbers given below. If you had more than one previous job and they differed in the amount of time spent on a task, use a code number that will best represent an overall average. (Do not be concerned that you have not accomplished many of the tasks listed as no student is expected to have performed all, or even most, of the tasks.)

CHECK (✓) ALL OF THE TASKS YOU HAVE PERFORMED IN THE ENTIRE COLUMN OF STEP 1 BEFORE YOU BEGIN TO ENTER THE TIME CODE FOR EACH IN STEP 2. ENTER TIME CODE ONLY FOR TASKS YOU HAVE PERFORMED.

TIME CODE

1. VERY MUCH BELOW AVERAGE TIME ON THIS TASK.
2. BELOW AVERAGE TIME.
3. AVERAGE TIME.
4. ABOVE AVERAGE TIME.
5. VERY MUCH ABOVE AVERAGE TIME ON THIS TASK.

(Time does not necessarily equal importance)

Example:	STEP 1 CHECK PREVIOUS JOB TASKS	STEP 2 ENTER TIME CODE
1. Issue moving traffic citations	✓	5
2. Issue warning tickets		
3. Administer field tests for intoxication (coordination tests, etc.)	✓	4
4. Check autos against stolen car list		
5. Recover stolen automobile	✓	2

A. POLICE OPERATIONS

	STEP 1 CHECK PREVIOUS JOB TASKS	STEP 2 ENTER TIME CODE
1. Issue moving traffic citations		
2. Issue warning tickets		
3. Administer field tests for intoxication (coordination tests, etc.)		
4. Check autos against stolen car list		
5. Recover stolen automobile		
6. Control spectator access to traffic crash scene		
7. Call for supplementary aid (e.g., wreckers, fire department) for traffic crashes		
8. Conduct preliminary investigation on hit and runs		
9. Prepare accident report		
10. Direct or control nonroutine congested traffic		
11. Report hazardous conditions		
12. Suggest changes in traffic control procedures		
13. Plan and direct selective traffic enforcement work		
14. Prepare accident statistical data		
15. Conduct special studies of accident records		
16. Calculate average accident rates		
17. Calculate roadway or intersection capacity		
18. Decide on the location for installing traffic control signs		
19. Advise city planners on traffic planning		
20. Advise city planners on traffic engineering needs		
21. Prepare officer's notebook		
22. Prepare research reports		
23. Create charts and graphs		

A. POLICE OPERATIONS (cont'd)

	STEP 1 CHECK PREVIOUS JOB TASKS	STEP 2 ENTER TIME CODE
24. Develop maps or charts		
25. Handle a family disturbance		
26. Handle an insane person		
27. Handle a bomb threat		
28. Guard a crime scene		
29. Review offense reports to establish M.O.		
30. Prepare missing person report		
31. Respond to riot		
32. Control spectators at civil disturbances		
33. Develop plan for crowd control operation		
34. Develop plan for organizing crowd control units or operations		
35. Implement restraining order against demonstrators		
36. Communicate with leaders of demonstrations		
37. Implement restraining order against strikers		
38. Communicate with labor and management over strike disturbances		
39. Develop disaster control plans		
40. Evaluate emergency procedures		
41. Maintain liaison with other law enforcement or other community agencies		
42. Coordinate police actions with other units or agencies		
43. Supervise law enforcement patrol tactics		
44. Supervise use of communications equipment or vehicles		
45. Supervise directing and controlling traffic		
46. Supervise investigation of traffic accidents		

A. POLICE OPERATION (cont'd)

	STEP 1 CHECK PREVIOUS JOB TASKS	STEP 2 ENTER TIME CODE
47. Supervise breakdown and study of traffic enforcement and accident data		
48. Supervise traffic case preparation		
49. Supervise criminal case preparation		
50. Supervise preliminary investigations		
51. Conduct weapons qualification activities		
52. Conduct proficiency or qualification training		
53. Instruct trainees		
54. Work as field training officer		
55. Participate in training conferences		
56. Conduct orientation of newly assigned personnel		
57. Conduct on-the-job training		
58. Interpret policies, directives, or procedures for subordinates		
59. Assign personnel to duty positions		
60. Plan or schedule work assignments or work loads		
61. Determine work priorities		
62. Analyze work load requirements		
63. Evaluate work schedules		
64. Develop or improve work methods or procedures		
65. Conduct driver education courses in public or private schools		
66. Provide traffic safety education to the public		
67. Solicit help from citizens in reporting and detecting crime		
68. Receive complaints on city services		
69. Refer citizens to appropriate private agency when their requests are not within the scope of police work		

B. CRIMINAL INVESTIGATION

	STEP 1 CHECK PREVIOUS JOB TASKS	STEP 2 ENTER TIME CODE
70. Conduct preliminary investigation of a larceny		
71. Conduct full investigation of a larceny		
72. Conduct preliminary investigation of an auto theft		
73. Conduct full investigation of an auto theft		
74. Recover stolen automobile		
75. Conduct preliminary investigation of a burglary		
76. Conduct full investigation of a burglary		
77. Conduct preliminary investigation of a robbery		
78. Conduct full investigation of a robbery		
79. Conduct preliminary investigation of an assault with deadly weapon		
80. Conduct full investigation of an assault with deadly weapon		
81. Conduct preliminary investigation of a murder		
82. Conduct full investigation of a murder		
83. Conduct preliminary investigation of a sudden or suspicious death		
84. Conduct full investigation of a sudden or suspicious death		
85. Conduct preliminary investigation of a fire		
86. Conduct full investigation of a fire		
87. Conduct preliminary investigation of a suspicious person		
88. Conduct full investigation of a suspicious person		
89. Conduct preliminary investigation of a rape		
90. Conduct full investigation of a rape		

B. CRIMINAL INVESTIGATION (cont'd)

	STEP 1 CHECK PREVIOUS JOB TASKS	STEP 2 ENTER TIME CODE
91. Conduct preliminary investigation of a sex offense other than rape		
92. Conduct full investigation of a sex offense other than rape		
93. Investigate narcotics violations		
94. Investigate gambling violations		
95. Investigate consumer complaints		
96. Investigate repossession complaints		
97. Solicit help from citizens in reporting and detecting crime		
98. Obtain "portrait parle" of suspect		
99. Conduct a criminal surveillance		
100. Question known offenders and persons who arouse suspicion		
101. Administer field tests for drug and alcohol use		
102. Make use of informants		
103. Arrest juvenile suspects		
104. Prepare affadavit for search warrant		
105. Conduct searches of buildings		
106. Collect physical evidence at scene of crime		
107. Label evidence collected at scene of crime		
108. Conduct the booking process of a suspect at police desk		
109. Check past record of prisoner for inclusion in reports		
110. Interrogate suspects		
111. Study background, rap sheet, and M.O. of suspects prior to interrogation		

B. CRIMINAL INVESTIGATION (cont'd)

		STEP 1 CHECK PREVIOUS JOB TASKS	STEP 2 ENTER TIME CODE
112.	Vary method of interrogation based on suspect's background, cultural differences		
113.	Advise prisoners of their legal and civil rights		
114.	Record identifying information obtained from prisoner		
115.	Interrogate suspects or witnesses with aid of polygraph		
116.	Operate breath analyzer		
117.	Operate voiceprinter		
118.	Operate spectrograph		
119.	Take photographs during investigation		
120.	Prepare arrest report		
121.	Prepare accident report		
122.	Prepare officer's notebook		
123.	Prepare missing person report		
124.	Prepare narrative report		
125.	Prepare suspect interrogation card		
126.	Prepare report on evidence collected		
127.	Prepare investigation report		
128.	Prepare juvenile investigation report		
129.	Prepare vice observation report		
130.	Prepare vehicle theft report		
131.	Review offense reports to establish M.O.		
132.	Supervise follow-up investigations		
133.	Direct investigation activities		
134.	Testify in court on criminal cases		
135.	Discuss criminal cases with prosecuter		

B. CRIMINAL INVESTIGATION (cont'd)

	STEP 1 CHECK PREVIOUS JOB TASKS	STEP 2 ENTER TIME CODE
136. Prepare criminal case summary sheet for prosecutor		
137. Prepare criminal case folders		
138. Prepare charge for magistrate		
139. Supervise criminal case preparation		

C. CRIMINAL AND CONSTITUTIONAL LAW

		STEP 1 CHECK PREVIOUS JOB TASKS	STEP 2 ENTER TIME CODE
140.	Prepare affadavit for search warrant		
141.	Conduct searches of buildings		
142.	Search vehicle for evidence and weapons		
143.	Search suspects for evidence and weapons		
144.	Collect physical evidence at scene of crime		
145.	Question known offenders and persons who arouse suspicion		
146.	Advise persons involved in minor infractions of alternatives to arrest		
147.	Advise prisoners of their legal and civil rights		
148.	Interrogate suspects		
149.	Prepare suspect interrogation card		
150.	Interrogate suspects or witnesses with aid of polygraph		
151.	Arrest juvenile suspects		
152.	Investigate consumer complaints		
153.	Obtain restraining order against demonstrators		
154.	Implement restraining order against demonstrators		
155.	Obtain restraining order against strikers		
156.	Implement restraining order against strikers		
157.	Communicate with labor and management over strike disturbances		

D. CRIMINAL EVIDENCE

	STEP 1 CHECK PREVIOUS JOB TASKS	STEP 2 ENTER TIME CODE
158. Collect physical evidence at scene of crime		
159. Label evidence collected at scene of crime		
160. Conduct searches of buildings		
161. Search vehicle for evidence and weapons		
162. Search suspect for evidence and weapons		
163. Investigate narcotics violations		
164. Investigate gambling violations		
165. Take photographs during investigation		
166. Administer field tests for drug and alcohol use		
167. Operate breath analyzer		
168. Operate voiceprinter		
169. Operate spectrograph		
170. Interrogate suspects or witnesses with aid of polygraph		

E. CRIMINAL PROCEDURES

	STEP 1 CHECK PREVIOUS JOB TASKS	STEP 2 ENTER TIME CODE
171. Prepare affadavit for search warrant		
172. Advise persons involved in minor infractions of alternatives to arrest		
173. Arrest juvenile suspects		
174. Conduct the booking process of a suspect at police desk		
175. Advise prisoners of their legal and civil rights		
176. Testify in court on criminal cases		
177. Discuss criminal cases with prosecutor		
178. Prepare charge for magistrate		
179. Prepare criminal case folders		
180. Prepare criminal case summary sheet for prosecutor		

F. ADMINISTRATION

	STEP 1 CHECK PREVIOUS JOB TASKS	STEP 2 ENTER TIME CODE
181. Locate, interview, and establish credibility of witnesses		
182. Take statements or depositions from witnesses		
183. Send evidence to lab for analysis		
184. Collect traffic crash evidence		
185. Interview those involved in traffic crash		
186. Determine point(s) of impact in traffic crash		
187. Determine key or crucial events related to the traffic crash		
188. Check crashed vehicles for evidence of mechanical defects that contributed to crash		
189. Photograph accident and/or crime scene		
190. Diagram and record measurement of traffic crash scene		
191. Prepare accident statistical data		
192. Conduct preliminary investigation on hit and runs		
193. Conduct follow-up on hit and runs		
194. Investigate complaints against police		
195. Explain goals and policies of police department to citizens		
196. Plan a human relations program		
197. Implement human relations programs		
198. Evaluate human relations programs		
199. Direct community relations program		
200. Evaluate individuals for promotion or demotion		
201. Evaluate and recommend personnel for special training or duty assignments		

F. ADMINISTRATION (cont'd)

	STEP 1 CHECK PREVIOUS JOB TASKS	STEP 2 ENTER TIME CODE
202. Evaluate compliance with work standards		
203. Establish performance standards		
204. Analyze work load requirements		
205. Evaluate work schedules		
206. Draft or revise job descriptions		
207. Evaluate job descriptions		
208. Develop organizational or functional charts		
209. Requisition supplies or equipment		
210. Maintain an inventory of supplies or equipment		
211. Direct maintenance of equipment		
212. Determine requirements for space, personnel equipment, or supplies		
213. Evaluate use of workspace, equipment, or supplies		
214. Plan layout of facilities		
215. Direct administrative police activities		
216. Direct maintenance of administrative files		
217. Evaluate administrative forms, files, or procedures		
218. Draft a budget		
219. Evaluate budget		
220. Establish organizational policies, office instructions, or special security instructions		
221. Maintain training records		
222. Plan a training program		
223. Implement training programs		
224. Direct training programs		

F. ADMINISTRATION (cont'd)

	STEP 1 CHECK PREVIOUS JOB TASKS	STEP 2 ENTER TIME CODE
225. Evaluate training programs		
226. Evaluate training methods, techniques, or programs		
227. Plan or evaluate on-the-job training programs		
228. Procure training aids, space, or equipment		
229. Participate in training conferences		
230. Conduct training conferences		
231. Determine training requirements		
232. Develop course curriculum materials		
233. Conduct orientation of newly assigned personnel		
234. Demonstrate how to locate technical information		
235. Plan or prepare briefings		
236. Conduct or participate in staff meetings		
237. Test, analyze, or evaluate police programs		
238. Evaluate emergency procedures		
239. Develop disaster control plans		
240. Develop plan for crowd control operations		
241. Plan a retention program		
242. Implement retention programs		
243. Evaluate retention programs		
244. Counsel personnel on personal or job-related problems		
245. Implement suggestion programs		
246. Implement safety programs		
247. Implement security programs		
248. Implement cost reduction programs		

F. ADMINISTRATION (cont'd)

	STEP 1 CHECK PREVIOUS JOB TASKS	STEP 2 ENTER TIME CODE
249. Plan or conduct inspection programs		
250. Make staff studies, surveys, or special reports		
251. Prepare research reports		
252. Create charts and graphs		
253. Draft directives, manuals, or supplements		
254. Direct analysis of reports		
255. Write press releases		
256. Prepare and present speeches		

V. GENERAL COMMENTS

List any other tasks that you performed on your previous positions in the law enforcement field that were not listed in the previous section. You may also make other general comments that you feel may be relevant to your work experience.

Guidelines and Procedures

Area D - Demonstration Validity Studies

Virtually all assessment involves expert judgment in one form or another. Thus, the demonstration of validity of expert judgment is validation in its most general sense and refers to the process of comparing evidence of learning obtained from different sources. Validation is a process which can be applied to almost any existing assessment procedure provided that data can be collected over a reasonable sample of students.

The same basic plan can be used as a guide by all institutions conducting demonstration validity studies. There are four major steps in conducting a validity study for Area D.

Step 1: Choose an appropriate assessment procedure.

An assessment procedure should be chosen which has at least two major characteristics. First, the procedure should be one for which it is possible to collect independent judgments or other concrete pieces of evidence which bear on the student's learning in the area to be assessed. An example would be an interview procedure in which one active interviewer and one nonparticipant observer both sit in on a series of assessment interviews and each make independent assessments of each student's learning.

A second essential feature of the assessment procedure chosen for a study is that enough data can be collected within the time period available (October 15, 1975 to February 29, 1976). Obviously, an assessment procedure which in the past has been applied to very few students should not be chosen for a study. It is highly desirable that data be collected for at least 20 students. Validity studies with samples below 20 will generally produce less trustworthy results.

In addition to these two major characteristics, institutions will probably want to consider other factors in choosing an assessment procedure such as convenience, cost, etc. It should be pointed out that it is by no means necessary to developing new assessment procedures for purposes of this study although that would be perfectly acceptable. Any assessment procedure including those that have been used for a period of several years may be studied.

Step 2: Write a detailed narrative description of the development of the assessment procedure.

The purpose of this step is to thoroughly document the rationale and methods used to arrive at the final assessment procedure to be studied. Attachment A suggests a series of questions which can serve as a guide for writing the narrative description.

Step 3: Collect data.

This is the most critical step in demonstrating the validity of an assessment procedure. The data that are to be collected can be in any of several forms including numerical ratings, number of credits awarded, or categorical judgments such as pass/fail, credit vs. no credit, etc. In order to be useful at least two independent assessments must be made for each student. In most cases, "independent" will mean that two different judges will, without collaboration, make an assessment of the same student after examining the same or similar evidence of learning.

> Example: Twenty students are assessed for learning in music through an actual piano performance. Each performance is rated by a judge present during the performance. A second judge without any knowledge of the first judge's rating also rates each performance after listening to a recording of each performance.

Step 4: Report results of data collection.

It is assumed that in most cases the independent judgments will be expressed in the form of a numerical rating for each student. Thus, for each of the 20 or more students sampled there should be two such (independent) estimates. For the assessment procedure to be considered valid there should be some consistency between these two sets of assessments.

The first step in reporting the results is to actually list the data for each of the 20 students. The students will, of course, be identified by code numbers rather than name. This will allow at least a subjective estimate of the consistency of the two independent assessments to be made. The second step in reporting the results of the data collection is to summarize the data in one or more ways. The following two methods are suggested: (a) frequency of agreement counts - data can also be summarized by reporting the number of times that judges agreed exactly, the number of times judges agreed within one scale point, category, or number of credit, the number of times judged within two scale points, etc., and (b) computing a correlation coefficient - a common statistical procedure for summarizing data of this type is to compute a correlation coefficient. The correlation coefficient is an index of a agreement which ranges from -1.0 to +1.0 where -1.0 would indicate "perfect" disagreement, 0 would indicate no consistency between judgments at all, and +1.0 would indicate perfect consistency. For the type of data presently considered correlation coefficients will generally fall in the range from 0 to +1.0. It is not essential to compute a correlation coefficient but for those participants wishing to do so computational procedures are given in Attachment C for the data reported in the example given in Attachment B.

It is assumed that participants will be thoroughly familiar with Working Paper No. 10. The contents of this working paper, particularly Chapter II presents some of the major concepts which are related to the validation of expert judgment. Other material may be helpful in structuring the particular assessment procedure chosen. In some cases, as with the interview presented in Chapter IV participants may want to adapt some materials for local use.

Schedule

The following are key dates for site coordinators in planning and executing the data collection.

Date	Event
Mid October	CAEL Assembly Meeting
Early November	Complete Plans for Data Collection
Mid November	Complete Meetings with Faculty Participants and Schedule
January 1	Complete at least 1/3 of the Data Collection in each of the Three Studies
January 8	Attend Meeting for Site Coordinators
March 1	Finish Data Collection and Mail to ETS

Attachment A

A Guide for Writing a Narrative Report of the Development of an Assessment Procedure

Listed below are a series of questions which can serve as a guide for participants documenting the development of the assessment procedure chosen for the demonstration validity study. These questions are by no means intended to constrain participants in the structure or extent of their report, but can serve as a fairly objective checklist for making sure the report includes the essential details.

1. How were criterion standards established (i.e., how is the evidence a student presents related to the number of credits or degree of competency awarded)?
    a. Who participated in establishing criterion standards (students, experts, measurement specialist)?
    b. How were criterion standards finally expressed? Where they expressed in terms of general definitions, relatively objective behavioral indicators, or some other method?
2. Why was the particular assessment procedure chosen?
    a. What factors were considered in choosing the specific assessment procedure?
3. What steps were involved in developing the assessment procedure?
    a. Who was involved?
    b. Was the assessment procedure first tried out and then revised?
    c. Did certain experts review the assessment procedure before it was applied?
4. Does the narrative fully describe the procedures used in conducting assessments?
    a. If there are any special provisions necessary for space, time, and equipment these should be detailed.
    b. What steps were taken to select and prepare assessors?
    c. What steps were taken to prepare students?

d. What provisions were made for recording the assessment either through note taking, rating scales, or by electronic means?

5. What procedures were used for collecting data? What provisions were made for ensuring independence of different judgments?

Note: Wherever possible, it would be desirable to supplement the narrative report with any materials developed for the purposes of assessment. Examples might include rating scales, instructions for assessors, instructions for students, etc.

Attachment B

Illustrative Example of Data Collected and Analyzed
For a Demonstration Validity Study

The data below are a hypothetical set of average ratings of artistic products submitted by 20 students for credit in art. Two experts assessed each student's work independently judging each one's work with respect to quality, depth, and breadth using the procedures outlined on pages 62 and 63 of CAEL Working Paper No. 10 to arrive at average ratings.

Student #	Judge A	Judge B
1	2	3
2	3	4
3	1	1
4	3	3
5	3	5
6	4	4
7	2	3
8	3	3
9	5	4
10	5	3
11	3	2
12	1	2
13	4	5
14	4	2
15	3	3
16	2	3
17	2	1
18	3	4
19	4	3
20	3	2

Example: Analysis of data by closeness of agreement.

Number of perfect agreements = 5

Number of ratings one point apart = 13

Number of ratings two points apart = 2

Attachment C

Computational Procedures for a Correlation Coefficient

In computing a correlation coefficient it is assumed that two sets of judgments or assessments are available for the sample of 20 students. The steps necessary to compute a correlation coefficient are presented below and in each case the computation is illustrated with the data from Attachment B.

Step	Result
A. Compute the sum of ratings for Judge A	A = 60
B. Compute the sum of ratings for Judge B	B = 60
C. Square each rating and compute the sum for Judge A	C = 204
D. Square each rating and compute the sum for Judge B	D = 204
E. For each student multiply the rating given by Judge A by the rating given by Judge B and compute the sum	E = 192
F. Calculate the first part of denominator as follows: first multiply the results of Step C by the number of students and then subtract the square of the results of Step A	$F = 20 \mathrm{x} C - A^2$ $= 20 \mathrm{x} 204 - 60^2$ $= 480$
G. Calculate the second part of the denominator by first multiplying the results of Step D by the number of students and subtracting the square of the results of Step B	$G = 20 \mathrm{x} C - B^2$ $= 20 \mathrm{x} 204 - 60^2$ $= 480$

Step	Result
H. Multiply the result of Step F by the result of Step G and then take the square root to obtain the denominator	$H = \sqrt{GxF}$ $= \sqrt{480x480}$ $= 480$
I. Calculate the numerator by first multiplying the result of Step E by the number of students and then subtracting the result of Step A multiplied by the result of Step B	$I = 20xE - AxB$ $= 20x192 - 60x60$ $= 240$
J. To obtain the correlation coefficient divide the numerator I by the denominator H	$J = I/H$ $= \frac{240}{480}$ $= .5$

Attachment A

A Guide for Writing a Narrative Report of the Development of an Assessment Procedure

Listed below are a series of questions which can serve as a guide for participants documenting the development of the assessment procedure chosen for the demonstration validity study. These questions are by no means intended to constrain participants in the structure or extent of their report, but can serve as a fairly objective checklist for making sure the report includes the essential details.

1. How were criterion standards established (i.e., how is the evidence a student presents related to the number of credits or degree of competency awarded)?
    a. Who participated in establishing criterion standards (students, experts, measurement specialist)?
    b. How were criterion standards finally expressed? Where they expressed in terms of general definitions, relatively objective behavioral indicators, or some other method?
2. Why was the particular assessment procedure chosen?
    a. What factors were considered in choosing the specific assessment procedure?
3. What steps were involved in developing the assessment procedure?
    a. Who was involved?
    b. Was the assessment procedure first tried out and then revised?
    c. Did certain experts review the assessment procedure before it was applied?
4. Does the narrative fully describe the procedures used in conducting assessments?
    a. If there are any special provisions necessary for space, time, and equipment these should be detailed.
    b. What steps were taken to select and prepare assessors?
    c. What steps were taken to prepare students?

   d. What provisions were made for recording the assessment either through note taking, rating scales, or by electronic means?

5. What procedures were used for collecting data? What provisions were made for ensuring independence of different judgments?

Note: Wherever possible, it would be desirable to supplement the narrative report with any materials developed for the purposes of assessment. Examples might include rating scales, instructions for assessors, instructions for students, etc.

Attachment B

Illustrative Example of Data Collected and Analyzed For a Demonstration Validity Study

The data below are a hypothetical set of average ratings of artistic products submitted by 20 students for credit in art. Two experts assessed each student's work independently judging each one's work with respect to quality, depth, and breadth using the procedures outlined on pages 62 and 63 of CAEL Working Paper No. 10 to arrive at average ratings.

Student #	Judge A	Judge B
1	2	3
2	3	4
3	1	1
4	3	3
5	3	5
6	4	4
7	2	3
8	3	3
9	5	4
10	5	3
11	3	2
12	1	2
13	4	5
14	4	2
15	3	3
16	2	3
17	2	1
18	3	4
19	4	3
20	3	2

Example: Analysis of data by closeness of agreement.

Number of perfect agreements = 5

Number of ratings one point apart = 13

Number of ratings two points apart = 2

Attachment C

Computational Procedures for a Correlation Coefficient

In computing a correlation coefficient it is assumed that two sets of judgments or assessments are available for the sample of 20 students. The steps necessary to compute a correlation coefficient are presented below and in each case the computation is illustrated with the data from Attachment B.

Step	Result
A. Compute the sum of ratings for Judge A	$A = 60$
B. Compute the sum of ratings for Judge B	$B = 60$
C. Square each rating and compute the sum for Judge A	$C = 204$
D. Square each rating and compute the sum for Judge B	$D = 204$
E. For each student multiply the rating given by Judge A by the rating given by Judge B and compute the sum	$E = 192$
F. Calculate the first part of denominator as follows: first multiply the results of Step C by the number of students and then subtract the square of the results of Step A	$F = 20xC - A^2$ $= 20x204 - 60^2$ $= 480$
G. Calculate the second part of the denominator by first multiplying the results of Step D by the number of students and subtracting the square of the results of Step B	$G = 20xC - B^2$ $= 20x204 - 60^2$ $= 480$

Step	Result
H. Multiply the result of Step F by the result of Step G and then take the square root to obtain the denominator	$H = \sqrt{GxF}$ $= \sqrt{480x480}$ $= 480$
I. Calculate the numerator by first multiplying the result of Step E by the number of students and then subtracting the result of Step A multiplied by the result of Step B	$I = 20xE - AxB$ $= 20x192 - 60x60$ $= 240$
J. To obtain the correlation coefficient divide the numerator I by the denominator H	$J = I/H$ $= \frac{240}{480}$ $= .5$

APPENDIX 5

Data Analysis

5a. Study A1
5b. Study A2
5c. Study A3
5d. Study B1
5e. Study B2
5f. Study B3
5g. Study B4
5h. Study C1
5i. Study C2
5j. Study C3

DATA ANALYSIS[*]

This appendix describes the research designs and statistical procedures used in analyzing the field research data. For each study, the analyses are presented in order of their occurrence within the text. Since most of the statistical procedures employed the analysis of variance, the description of each analysis typically includes the following:

1. the independent and dependent variables
2. the experimental design
3. the ANOVA model (using the notation of Winer, 1971), if appropriate
4. the purpose of the analysis
5. special problems, if any
6. summary table of the analysis results or illustration of a typical analysis.

Throughout the text, reliability coefficients are presented to depict the degree of relationship in a set of ratings. For example, it is useful to summarize as a single coefficient the consistency among three credit awards made to students by three faculty assessors. Traditional formulations of reliability represent the reliability coefficient as the ratio of the variance of "true" scores to the variance of observed scores. When using the analysis of variance model, one of several versions of the intraclass correlation is used to represent this ratio. The formula which is used here,

$$r_{(\text{Intraclass})} = \frac{\text{MS(Persons)} - \text{MS(Error)}}{\text{MS(Persons)}}$$

---

[*] This section was written by Kurt F. Geisinger in consultation with Richard R. Reilly and Warren W. Willingham who were largely responsible for the original design of the studies.

is analogous to the traditional coefficient alpha. As such, this coefficient indexes the internal consistency of the ratings; like coefficient alpha, it is the average correlation between all possible "split-halves" of the unit in question. In the numerator, the mean square for persons (or portfolios, learning instances, etc.) minus the mean square for residual or error (typically represented as either between-rater, within-cell variation, or person-by-condition interaction) represents "true" variation. Clearly, this coefficient has a maximum value of one. While negative values are computationally possible, such values are due to sampling fluctuations of the mean square terms involved; in the text of this report, such values are reported as zero.

In estimating interrater agreement, the rater main effect is associated with "rater-error." Depending upon the nature of the operational notation, such variation may be considered part of the error component or it may be ignored. For example, if judgments are normalized by judge, this source of variation should be ignored. In crediting experiential learning it seems generally appropriate to assume that discrepancies among average ratings of different judges do constitute a source of inconsistency in credit practices. Consequently, interrater reliability estimates are here based upon an ANOVA model that treats ratings from different judges as random observations. In several instances, as noted in text, the rater main effect is ignored in order to gauge the extent to which interrater reliability might be improved by reducing the rater main effect through training. These two types of interrater reliability are here called "interrater agreement" and "interrater consistency" (the latter disregarding the rater as a source of error--see Page 19 of Chapter III).

Eta values are presented along with many of the ANOVAs in this appendix. Such etas serve as indices of the explanatory power of a main effect, an interaction, or some combination of effects in an ANOVA. Unlike F-values, etas are not subject to differential interpretation dependent upon the number of cases involved in the analysis. Furthermore, etas include both systematic linear and curvilinear variance in characterizing the effect in question. The definitional formula for eta follows:

$$(\text{eta})^2 = 1 - \text{Relative Error.}$$

The formula used in this report for the calculation of etas for main effects follows:

$$(\text{eta})^2 = \frac{\text{Sum of Squares for Main Effect}}{\text{Total Sum of Squares}}$$

and then,

$$\text{eta} = \sqrt{(\text{eta})^2}$$

More than one source of variance can be combined by simply summing their respective sums of squares in the numerator and then dividing by the total sum of squares for the design. For example, if one wishes to attribute both a Rater Main Effect and the Rater-by-Subject Interaction as Rater-Related Variance, one could follow the above procedure. Eta coefficients do have a drawback. Due to overfitting, they tend to be related to the number of degrees of freedom associated with the term being evaluated. This is especially true in calculating eta coefficients for interaction terms, where degrees of freedom can be comparatively large. As a result, interaction effects which are weak can appear spuriously large. For this reason, only eta coefficients representing main effect terms are presented; the strength of interaction terms may only be estimated from the statistical significance test.

The following references were used extensively and should be consulted for additional information: Cronbach, L. J., Gleser, G. C., Nanda, H., and Rajaratnam, N.; Ebel, R. L.; Kim, J. O., and Kohout, F. J.; Lord, F. M., and Novick, M. R.; Nunnally, J. C.; Wiggins, J. S., and Winer, B. J.

Appendix 5a: Study A1

1. The first set of analyses in A1 were two-way, fixed effects analyses of variance (ANOVAs), using institutions as one factor and illustrative cases as the second factor. Three such analyses were performed, one each for psychology, management, and education. The ANOVA model is:

$$X_{ijk} = \mu_{..} + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk} ,$$

where X_{ijk}, the dependent variable, is total semester credits awarded an illustrative case by a faculty rater; α_i and β_j represent the main effects of cases and institutions, respectively; $\alpha\beta_{ij}$ denotes their interaction, and error is the pooled estimate of within-institution faculty rating differences. In management and education, data from nine institutions were analyzed; due to missing data, only eight institutions were represented in psychology. Most illustrative cases were rated by three faculty members per institution, but in some cases, only two such ratings were made. The purpose of this analysis was two-fold: to estimate the extent of institutional differences in average credit awarded and to investigate whether institutions assessed the illustrative cases differently (that is, whether the case x institution interaction was significant). The answer to the first of these questions is noted on pp. 10-12 of Chapter IV. In answer to the second, the case x institution interaction is significant for psychology, but not for either management or education. The following source table for education exemplifies these analyses:

Source	Sum of Squares	df	Mean Square	F-Value
Institution	23,400.9	8	2925.1	4.81**
Case	3,695.7	4	923.9	1.52
Institution x Case	7,085.3	32	221.4	0.29
Error	58,645.4	76	771.7	
Total	92,416.7	120		

Eta(Institution) = .503
**p<.01

2. A second analysis in A1 was performed to determine whether the three content areas awarded credit differently across institutions. This analysis is referred to under Question 2. The analysis is a two-way fixed effects ANOVA with Institutions (eight levels) and Content Areas (three levels) as main effects. The dependent variable is the average number of credits (averaged over faculty raters within each institution) awarded an illustrative case. The model for this analysis is identical to the previous analysis:

$$X_{ijk} = \mu_{..} + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk} ,$$

except that here, α_i represents the content area main effect. The five illustrative cases serve as the within-cell observations. The summary table for the analysis follows:

Source	Sum of Squares	df	Mean Square	F-Value
Institution	420,090.98	7	60,129.00	3.28**
Content Area	341,428.74	2	170,714.37	9.31**
Institution x Content	881,045.96	96	62,931.83	3.43**
Error	1,760,763.31	119		

Eta(Institution) = .488
Eta(Content Area) = .440
**p<.01

3. A third A1 analysis is used to help investigate the extent of faculty agreement mentioned in Question 3. For each illustrative case, the number of credits awarded by each faculty member within an institution is subtracted from the number of credits assigned it by every other faculty rater within that same institution. For instance, if one illustrative case is rated by three faculty members within an institution, three absolute differences can be computed: between raters one and two, between raters one and three, and between raters two and three. Then, these distributions of absolute differences are tabulated over institutions to form an overall frequency distribution of differences within institutions in rated credits within a content area. Three of these frequency distributions have been developed (Table A1-3), one for each content area. Finally, these three content area-specific distributions were summed to generate an overall frequency distribution.

4. The fourth and final A1 analysis consisted of 27 one-way ANOVAs. Each ANOVA was institution and content area-specific and used illustrative cases as the fixed main effect. The dependent variable was the total

number of credits awarded each case by each faculty rater. The model for the analysis within each institution is:

$$X_{ij} = \mu. + \alpha_i + \varepsilon_{ij} ,$$

where X_{ij} is the credit assigned the illustrative case by a rater; α_i is the "true" individual case differences, and the error consists of between-rater differences. The values assigned by each rater to a particular case are treated as within-cell observations. The coefficients generated estimate the internal consistency of the sum of the ratings, and were "stepped down" to two raters via the Spearman-Brown formula. All negative results were reported as zero. Lord and Novick (Chapters 7-8, 1968) discuss fully the rationale why such coefficients should be interpreted as zero. An illustrative analysis follows:

Source	Sum of Squares	df	Mean Square	F-Value
Cases	34,693.33	4	8,673.33	5.78
Error	15,000.00	10	1,500.00	
Total	49,693.33	14		

$$\text{Reliability (sum of 3 raters)} = 1 - \frac{\text{MS(Error)}}{\text{MS(Cases)}} = 1 - \frac{1500.00}{8673.33} = .83$$

$$\text{Reliability (sum of 2 raters)} = \frac{.667 \times .83}{1 + (.667-1.0).83} = .76 \ .$$

Appendix 5b: Study A2

1. The first A2 analysis aimed at determining the equivalency of ratings across the two methods: interview and role-playing. A second research question is to determine the extent of institutional differences. Specifically, this analysis compared the institutional means for each institution across the two methods. This difference between methods is referred to in response to Question 4; Table A2-2 summarizes the means on each method for all the institutions. The dependent variable in this analysis was average overall rating (that is, the sum of two overall ratings from two faculty judges, divided by two) for each student. Students were considered as randomly occurring within-cell units in a two-way ANOVA. The first factor in the ANOVA was the Institutional factor. The second factor was considered a repeated measure; this was the "Method" factor (Role-Playing versus Interviewing). The ANOVA model for this analysis is the fixed model, described in the following manner:

$$X_{ij} = \mu_{..} + \pi_i + \tau_j + \pi\tau_{ij} + \varepsilon_{ij} \quad .$$

Here, τ_j corresponds to the fixed "Methods" factor, and π_i corresponds to the random Institution factor. The interaction term, $\pi\tau_{ij}$, and the error term, ε_{ij}, are also both considered to be random variables. Notice in the summary table of the analysis presented below, that the mean square for interaction is used as the error term against which this "Methods" factor is tested. Both the Institutions factor and the Interaction effect are tested against the error term.

Source	Sum of Squares	df	Mean Square	F-Ratio
Institutions	65.47	7	9.35	3.43**
Methods	8.80	1	8.80	3.68
Interaction	16.73	7	2.39	.88
Error	526.67	198	2.73	
Total	617.56	208		

Eta(Institutions) = .326
Eta(Methods) = .119
$^{**}p<.01$

2. A second computer analysis was made to construct Table A2-3. The two interview ratings for each student were summed. This was also done for each student on role-playing. Using these two sums per student, a ten-by-ten cross-frequency distribution was formed. The numbers within the cells represent the number of students with scores on the corresponding methods' marginals. The table was presented to display the degree of relationship between the two methods.

3. A third analysis used was the construction of a sixteen-by-sixteen correlation matrix. The sixteen variables are comprised of the seven rating factors on "Traits" for both methods and the two overall ratings, one for each method. Again, the two faculty ratings were averaged. Several of these correlations are mentioned in the text in response to Question 5 and in Table A2-5.

4. A pair of three-factor ANOVAs comprise the fourth A2 analyses. These analyses are relevant to research Question 6, as discussed in the text, and the means are presented in Table A2-4. For each method a three-factor ANOVA was performed, using the following model:

$$\overline{ABC}_{ij} = \mu_{\cdot\cdot\cdot} + \alpha_i + \beta_{j(i)} + \gamma_k + \alpha\gamma_{ik} + \beta\gamma_{j(i)k} + \bar{\varepsilon}_{ijk} \quad .$$

In this model the α_i term refers to the first factor, Institutions. The second factor, $\beta_{j(i)}$, refers to the Students factor, which is nested within institutions. The final factor, γ_k, is a repeated measure, the seven counseling skill traits on factors. The first interaction term, $\alpha\gamma_{ik}$, refers to the interaction of institutions and traits. The second interaction term, $\beta\gamma_{j(i)k}$, represents a pooling of the student by trait interaction over all institutions. The error term, $\bar{\varepsilon}_{ijk}$, is based on the two faculty ratings of each student on each trait. There are several purposes for this analysis. A significant "Trait" main effect would represent unequal means for the seven factors. Again, the extent of institutional differences can be perceived from the Institution factor. The Student-by-Trait interaction will be high (and significant) if students are rated differentially according to their particular characteristics by faculty assessors; likewise, if this interaction term is low, students are not being differentially described and a less cumbersome rating system should be imposed. Of final interest is the school-by-trait interaction. A significant interaction would suggest rating patterns characteristic at particular institutions through ratings made by faculty assessors; the interaction is confounded with student differences across institutions. Both summary tables follow:

Interview Summary Table

Source	Sum of Squares	df	Mean Square	F-Ratio
Between SS	1170.8	137	8.55	
Institution	265.9	8	33.23	4.74**
Student/Institution	904.9	129	7.01	13.41**
Within Students	379.8	813	0.47	
Traits	45.8	6	7.63	21.22**
Traits x Institution	61.2	48	1.28	3.55**
Traits x Student/Institution	272.8	759	0.36	0.69
Residual	433.5	829	0.52	
Total	1984.1	1779		

Eta(Institution) = .366
Eta(Traits) = .152
** p<.01

Role-Playing Summary Table

Source	Sum of Squares	df	Mean Square	F-Ratio
Between SS	1099.9	137	8.03	
Institution	219.6	8	27.45	4.02**
Student/Institution	880.3	129	6.82	12.88**
Within Student	345.4	823	0.42	
Traits	45.4	6	7.57	22.42**
Traits x Institution	40.4	48	0.84	2.50**
Traits x Student/Institution	259.6	769	0.34	0.64
Residual	443.0	836	0.53	
Total	1888.3	1796		

Eta(Institution) = .341
Eta(Traits) = .155
** p<.01

5. Research Question 7 requires one-way ANOVAs to estimate the intraclass reliabilities of assessment for each method at each institution. The students within each institution represent levels of the main factor. Their two ratings (method of assessment-specific) are treated as randomly assigned observations; these are the dependent variables. The ANOVA model is:

$$X_{ij} = \mu. + \pi_i + \varepsilon_{ij} \quad ,$$

where π_i represents the student effect and ε_{ij} the error due to between-rater differences. As an example, the summary table and resultant intraclass calculations for Institution B's role-playing data follows. (The coefficients for all institutions under each method are presented in Table A2-6).

Institution B: Role-Playing

Source	Sum of Squares	df	Mean Square
Students	24.87	14	1.78
Error	8.50	15	0.57
Total	33.37	29	

$$\text{Reliability} = 1 - \frac{\text{MS(Error)}}{\text{MS(Students)}} = 1 - \frac{0.57}{1.78} = .68$$

6. Another analysis was performed for each institution to show the relationship between the scores assigned by two different methods. Here, the two ratings (made by two assessors) for each method were summed. Then, these two sums were correlated. That is, the total score (based on two ratings) each person within an institution made under the interview

technique was correlated with the score made under the role-playing technique. These coefficients are found in Table A2-6 and are comparable to coefficients of construct (or convergent) validity.

7. A final analysis was performed, primarily to look at the extent of student-by-method interaction. If such an interaction were significant, then some students would receive significantly more credit when assessed by one method; other students would receive significantly more when assessed using the other methods. This analysis was constructed as a three-factor ANOVA. The Students factor was nested within levels of the Institution factor. Both of these factors were then crossed with the Method of Assessment factor, a repeated measure. The ANOVA model for this analysis is:

$$\overline{ABC}_{ijk} = \mu_{..} + \alpha_i + \beta_{j(i)} + \gamma_k + \alpha\gamma_{ik} + \beta\gamma_{j(i)k} + \bar{\varepsilon}_{ijk} \ .$$

This model is identical to that of the fourth analysis above, except that the third factor, γ_k , refers to method of assessment, rather than trait. The two interactions are changed accordingly. The dependent variable for this analysis is overall rating; the two ratings (from different faculty assessors) are treated as randomly assigned within-cell observations, and constitute the $\bar{\varepsilon}_{ijk}$ term. The summary table for this analysis follows:

Source	Sum of Squares	df	Mean Square	F-Ratio
Between SS	316.7	142	2.23	
Institution	49.7	8	6.21	3.12**
Student/Institution	267.0	134	1.99	4.72**
Within Student	83.8	133	0.63	
Method	1.8	1	1.77	3.07
Method x Institution	11.1	8	1.39	2.42*
Method x Student/Institution	71.0	124	0.57	1.36*
Residual	101.0	239	0.42	
Total	501.5	514		

Eta(Institution) = .315
Eta(Method) = .060
**p<.01
*p<.05

Appendix 5c: Study A3

1. The first A3 analysis estimates the interjudge agreement in rating of students' responses. If the rating process is not reliably conducted, this data is of little value. The second analysis in A3 is the more important analysis. It is, however, contingent upon satisfactory reliability in the first analysis. This first analysis is a one-way analysis of variance with students serving as levels of the main factor and the four judgments (each made independently by an expert rater) serving as within-cell observations. The model for this analysis is:

$$X_{ijk} = \mu. + \alpha_i + \varepsilon_{ij} \quad ,$$

where X_{ijk} , the dependent variable, is the jth expert's rating of the ith student's response. The α_i effect concerns individual differences between students. The error term is composed of differences between raters. It should be noted that in this analysis, mean differences over students by the expert judges are attributed to error. The summary table for the analysis is as follows:

Source	Sum of Squares	df	Mean Square	F-Ratio
Students	167.08	110	1.52	5.47**
Error	92.50	333	0.28	
Total	259.58	443		

** $p<.01$

In this case, reliability was estimated as:

$$1 = \frac{MS(Error)}{MS(Students)} = 1 - \frac{0.28}{1.52} = .82 \ .$$

2. A second A3 analysis evaluates mean differences between the pre-assessment and post-assessment students. It must be remembered in this study that the same individuals were not tested twice; rather, two distinct groups were used from each institution. Scores illustrative of this design are found in Table A3-1; the means and the number of subjects from each institution are presented in Table A3-2. Behind the discussion of this research (Question 9) is a two-way fixed effects ANOVA. The first factor is the eight-level institution factor. The second factor is the dichotomous status of pre-assessment or post-assessment. The ANOVA model for the analysis is:

$$X_{ijk} = \mu_{..} + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk} \ .$$

Where X_{ijk}, the dependent variable, equals the sum of the four overall ratings assigned each student. The first effect term, α_i, represents the institution effect and the second, β_j, represents the pre-assessment versus post-assessment effect. Their interaction, $\alpha\beta_{ij}$, is also tested. Students are treated as random replications within each institutional assessment status cell. The within-cell variability equals error. The summary table follows:

Source	Sum of Squares	df	Mean Square	F-Ratio
Institution	310.9	7	44.4	9.76**
Pre- versus Post-Assessment	3.5	1	3.5	0.77
Main Effect Interaction	21.8	7	3.1	0.68
Error	623.4	137	4.6	
Total***	1172.7	152		

Eta(Institution) = .515
Eta(Pre-Post) = .055
**p<.01
***Note that due to the unbalanced nature of the design, the orthogonality of factors is lost and the sum of squares is no longer additive. The F-ratio and the resultant significance test is reasonably accurate, however.

Appendix 5d: Study B1

1. The third research question in Area B requires two ANOVAs using the same two sets of institutions as in the fourth analysis: Institutions A, B, and C, and Institutions H, I, and J. These two groups of institutions were analyzed separately since the slightly different portfolios used by the two sets may have altered judgments regarding the credit awards. The question of whether institutions evaluate portfolios differently can be stated statistically as, "Is there a significant institution by illustrative case interaction?" These two analyses were set up as two-factor fixed effect ANOVAs, with institutions and cases as the main factors. The ANOVA model follows:

$$X_{ijk} = \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk} ,$$

where X_{ijk} the dependent variable is the number of credits assigned as illustrative case by faculty assessor. The α_i effect represents individual case differences (fourteen levels); the β_j effect (three levels) represents institutional differences in average amount of credit awarded. The interaction term, $\alpha\beta_{ij}$, represents differences in the ranking of illustrative cases across institutions in terms of credit worthiness. The two faculty ratings per case within each institution represent within-cell observations (error). In the case of both analyses performed, that for A, B, and C, and H, I, and J, both main factors and their interaction are statistically significant. The summary table for the Institutions A, B, and C analysis follows:

Institutions A, B, C

Source	Sum of Squares	df	Mean Square	F-Ratio
Cases	19,501.9	13	1,500.14	12.00**
Institution	11,610.2	2	5,805.00	46.50**
Case x Institution	10,611.2	26	408.10	3.27**
Error	4,998.5	40	125.00	
Total	48,844.6	81		

Eta(Institution) = .488
Eta(Case x Institution) = .466
**p<.01

2. Table B1-2 and its related textual discussion also shows institutional reliability coefficients. These intraclass coefficients are calculated from mean squares arising from one-way ANOVAs (for each institution). The levels of the factor are students (for the local portfolios) or illustrative cases. The within-cell observations are the pairs of faculty assessments. The dependent variable is the number of credits awarded. The ANOVA model for this analysis is:

$$X_{ij} = \mu. + \pi_i + \varepsilon_{ij} ,$$

where π_i refers to individual differences (true score), and the error term ε_{ij} is based between-rater discrepancies. The institutional means represent the average number of credits awarded both illustrative and local portfolios. The averages are discussed under Question 1 and reported in Table B1-2. There is a minor problem in the averaging procedure. There is an occasional missing piece of data (such as only one rating of an illustrative portfolio rather than two). Hence, the means across institutions are not absolutely comparable. An illustrative summary table follows, analyzed using Institution H's illustrative portfolio data.

Institution H, Illustrative Portfolio Data

Source	Sum of Squares	df	Mean Square
Cases	856.5	13	65.88
Error	225.5	11	20.50
Total	1,378.0	24	

Notice that because there is some missing data in this analysis, the sums of squares for cases and error do not sum to the total. This also means that true and error scores are not completely orthogonal. The coefficients presented in Table B1-2 are coefficients of internal consistency based on two raters, they are calculated in the following manner.

$$\text{Reliability} = 1 - \frac{\text{MS(Error)}}{\text{MS(Cases)}} = 1 - \frac{20.50}{65.88} = 1 - .31 = .69 \quad .$$

3. In order to make a more global statement concerning the reliability of local portfolio assessment than the institution-specific coefficients, a pooled estimate of local portfolio assessment reliability was calculated. This coefficient was estimated using data from the previous one-way ANOVAs. The sum of squares (Students), sum of squares (Error), degrees of freedom (Students), and degrees of freedom (Error) were all summed over institutions. This procedure treats the data as nested ANOVA design. The resultant values are as follows:

Sum of squares (Students) = 31,795.92

Sum of squares (Error) = 6,669.00

Degrees of freedom (Students) = 87

Degrees of freedom (Error) = 93

From these values, mean square values for students and error can be calculated, resulting in values of 365.47 and 71.71, respectively. The pooled estimate of reliability is then calculated in the usual manner:

$$1 - \frac{\text{MS(Error)}}{\text{MS(Students)}} = 1 - \frac{71.71}{365.47} = .804 \quad .$$

This compares with a median reliability value of .76, as reported in Table B1-2.

4. A fourth analysis is referred to in the discussion of the first question. This analysis was based upon two sets of institutions: Institutions A, B, and C, which used one set of identical illustrative portfolios, and Institutions H, I, and J, which used a slightly different set of illustrative portfolios (see related sections in the text). Because these institutions used identical illustrative portfolios, it is possible to compare faculty assessments directly. This analysis computed differences (in number of credits awarded) between all faculty ratings on the same illustrative portfolios. Then, these differences were divided into two groups of differences as reported in Table B1-4. All differences which arise from faculty members of the same academic institution form the frequency distribution called "within-institution rater differences." All differences between-raters of different institutions form the frequency distribution called "between-institution rater differences." For instance, if the number of credits Professor X from University A assigned Portfolio 1 is subtracted from that Professor Y from University A assigned Portfolio 1,

a within-institution rater difference occurs. On the other hand, if the number of credits Professor X above assigned Portfolio 1 was subtracted from the number Professor Z at University C assigned Portfolio 1, a between-institution rater difference would occur. Comparing these distributions clearly shows the extent of institutional differences in crediting portfolios.

Appendix 5e: Study B2

1. This first analysis concerns Question 4 and provides reliability coefficients for each institution (see Table B2-2). These coefficients are based on data from five faculty ratings. First, a one-way ANOVA was computed where learning instances represent levels of the factor, the five ratings represent randomly assigned within-cell observations, and the dependent variable is the rating (0-3). The ANOVA model for such a design is:

$$X_{ij} = \mu. + \pi_i + \varepsilon_{ij} \ ,$$

where X_{ij} represents the dependent variable; π_i , the learning instance (true score); and ε_{ij} , the difference between raters. An exemplary summary table, using data from Institution E, follows.

Institution E ANOVA

Source	Sum of Squares	df	Mean Square	F-Ratio
Learning Instances	22.4	14	1.60	5.45**
Error	17.6	60	.29	
Total	40.0	74		

** $p<.01$

The mean square values for Learning Instances and Error are taken from the summary table above to form the reliability coefficient based on ratings summed over five faculty assessors. This is done below, again for data from Institution E.

$$\text{Reliability (5 raters)} = 1 - \frac{\text{MS(Error)}}{\text{MS(Instances)}} = 1 - \frac{.29}{1.6} = .817 \ .$$

However, to remain consistent with other coefficients within The Validation Report, the coefficients are then "stepped down" to two raters as follows:

$$\text{Reliability (2 raters)} = \frac{.4\ [\text{Reliability (5 raters)}]}{1 + (.4-1)[\text{Reliability (5 raters)}]} =$$

$$\frac{(.4)(.817)}{1 + (-.6)(.817)} = .64 \quad .$$

2. The second B2 analysis is aimed at both research Questions 5 and 6. The fifth research question requires examination of average ratings calculated over all instances in each institution. These means are calculated as by-products (actually, as marginal means) in a two-way ANOVA used to answer the sixth research question. The two factors in this ANOVA design are Institutions (10 levels) and Learning Instances (15 levels). The main question to be resolved concerns the significance of the Institution main effect and the Institution-by-Learning Instance interaction. As mentioned in the discussion in the text, simple inspection of the means of the Learning Instances shows large differences. Determination of the statistical significance of these differences is the second purpose for the analysis. The ANOVA model is:

$$X_{ijk} = \mu_{..} + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk} \quad ,$$

where X_{ijk} refers to the rating assigned by a faculty member at a specific institution to a specific learning instance. The α_i , β_j , and $\alpha\beta_{ij}$ terms refer to the Institution effect, the Learning Instance effect, and their interaction. The error term represents differences between the five raters from each institution. The summary table of the analysis follows:

Source	Sum of Squares	df	Mean Square	F-Ratio
Institutions	34.75	9	3.86	7.53**
Learning Instances	179.78	14	12.84	25.05**
Interaction	62.85	126	0.50	0.97
Error	302.46	590	0.51	
Total	579.85	739		

Eta(Institution) = .245

**$p < .01$

Appendix 5f: Study B3

1. The first B3 analysis provides the reliability coefficients as presented in Table B3-4. These coefficients give an indication of the extent of agreement within institutions concerning the requirement of specific types of documentation. Each of these coefficients is an intraclass coefficient calculated using mean square terms from one-way ANOVAs. Each one-way ANOVA is specific both to an institution and a type of documentation. The ANOVA model for these ANOVAs is:

$$X_{ij} = \mu. + \pi_i + \varepsilon_{ij} \quad .$$

The dependent variable, X_{ij} , refers to a one or a zero, depending on whether a faculty member from the institution in question required the type of documentation. The main effect, π_i , refers to (true) differences between the learning examples. An illustrative summary table, taken from Institution H's ratings concerns requiring "Letters" as a form of documentation.

Institution H, Type of Documentation: Letters

Source	Sum of Squares	df	Mean Square	F-Ratio
Learning Instances	5.28	14	0.377	4.71
Error	4.80	60	0.080	
Total	9.08	74		

Now, the intraclass coefficient, $1 - \frac{\text{MS(Error)}}{\text{MS(Examples)}}$ can be used to yield a coefficient of internal consistency between the five raters above. In this case, this coefficient becomes: $1 - \frac{0.080}{0.377} = 0.7879$. The coefficient computed above is based on five raters; within The Validation Report all such coefficients are being reported to two raters. The preceding coefficient must, therefore, be "stepped down" using the Spearman-Brown formula:

$$\text{Reliability (2 raters)} = \frac{.4\ (.7879)}{1 + (-.6)(.7879)} = .69 \quad .$$

2. The second B3 analysis is a three-way, fixed effects ANOVA. The three completely crossed factors are Institutions, Categories of Documentation, and Illustrative Cases, respectively. The ANOVA model for this analysis is:

$$X_{ijkl} = \mu_{...} + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk} + \varepsilon_{ijkl} \ .$$

The dependent variable, X_{ijkl} , is the dichotomous decision (0 or 1) as to whether a faculty member from a particular academic institution requires an illustrative case to present one of the fifteen types of documentation (listed in the Description of Study section) or not. Error is defined as the differences between faculty raters from the same academic institution. The values presented in Table B3-2 are row means taken from this analysis. (These proportions are summed over institutions.) The tabled proportions are simply the number of faculty members who required a type of documentation divided by fifty. The value in this form of data presentation is that certain characteristics of the data are striking. Letters are required most frequently; professional recognition is not commonly required. The data in this table respond to research Question 7 reasonably clearly. The values presented in Table B3-3 are also row means. For each pairing of type of documentation and institution, the proportion of times the type of documentation is required is summed over the illustrative cases (examples of learning). Since there are five raters per institution and fifteen examples of learning, there are seventy-five possible instances in which an institution could require a specific type of documentation. The values presented in Table B3-3 are simply the number of times the type of documentation is required by an institution divided by seventy-five. The summary table for this analysis is not referenced in the text; however, it is presented here.

Three Factor ANOVA

Source	Sum of Squares	df	Mean Square	F-Ratio
Institution	12.3	9	1.36	7.84**
Sample of Learning	18.5	14	1.32	7.60**
Type of Documentation	127.1	3	42.36	243.94**
Institution x Example of Learning	14.3	126	0.11	0.65
Institution x Type of Documentation	8.0	27	0.30	1.72*
Example of Learning x Type of Documentation	88.3	42	2.10	12.11**
Three-way Interaction	48.2	378	0.13	0.74
Error	416.8	2400	0.17	
Total	733.5	2999		

Eta(Institution) = .129
Eta(Category) = .159
**p<.01
*p<.05

3. The discussion following the presentation of research Question 8 mentions that there are statistically significant institutional differences in requiring all types of documentation. To test institutional differences in demanding documentation for crediting learning, four two-way ANOVAs were used; an ANOVA for each type of documentation was performed. The ten-level, fixed-effect Institution factor was crossed with the fifteen Examples of Learning. The ANOVA model for this analysis is:

$$X_{ijk} = \mu_{..} + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk} ,$$

where X_{ijk}, the dependent variable, represents a zero or a one made by a faculty assessor from a particular academic institution on a particular learning example. The α_i effect refers to institutional differences; the β_j effect corresponds to learning example differences. As noted, the model allows for an interaction of the main effects and the error term, ε_{ijk}, consists of between-rater differences which occur within a cell; that is, within an institution. In all four cases, Letters, Work Descriptions, Professional Recognition, and Work Record, both main effects were significant and the interaction was not significant. The analysis for Letters, which appears representative of the four, follows:

Type of Documentation: Letters

Source	Sum of Squares	df	Mean Square	F-Ratio
Institution	3.952	9	0.439	3.33**
Learning Instances	33.899	14	2.421	18.34**
Interaction	11.088	126	0.088	0.67
Error	79.199	600	0.132	
Total	128.138	749		

** $p<.01$

Appendix 5g: Study B4

1. The first B4 analysis is necessary, but not important in itself; but estimation of the reliability of student questionnaire ratings bears upon interpretation of B4 research questions. Therefore, this first B4 analysis estimates the reliability of the sum of the four expert raters' assessments of the student questionnaires. A one-way ANOVA design was used for this analysis. The 165 student questionnaires serve as levels of the factor; the four ratings are treated as randomly assigned within-cell observations. (There were five ratings left blank and treated as missing data.) The ANOVA model for this analysis follows:

$$X_{ij} = \mu. + \pi_i + \varepsilon_{ij} ,$$

where X_{ij} equals the dependent variable, a rating made of a student's response by a faculty assessor. Student differences (average over raters) are considered as the π_i effect, or true score. On the other hand, differences in ratings of the same student between raters is considered error, here represented by ε_{ij}. The summary table and resultant calculations of the intraclass reliability coefficient follow:

Source	Sum of Squares	df	Mean Square	F-Ratio
Students	217.9	164	1.33	3.85**
Error	169.1	490	.35	
Total	387.0	654		

** $p<.01$

$$\text{Reliability} = 1 - \frac{\text{MS(Error)}}{\text{MS(Students)}} = 1 - \frac{(\ .35)}{(1.33)} = .74 \ .$$

2. The second B4 analysis is also preliminary to the matrix analyses. The question is whether ratings of students' self-appraisals indicate the ability of the raters to differentiate responses of students, question-by-question in an individual fashion. Statistically, this is indicated by a significant student-by-question interaction. For example, if this interaction is not significant, it is more sensible to consider each student by only one value which corresponds to their general self-awareness. If the interaction is significant, there is some meaning in the five separate scales; each adds some unique information. As is implied above, the analysis is a two-way fixed-effects ANOVA, with the 165 students as levels of the Students' factor, and the five questions as levels of the Questions' factor. The ANOVA model for this analysis is:

$$X_{ijk} = \mu_{...} + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk} \quad ,$$

where X_{ijk} is the rating an expert assessor assigns a particular student on a particular question. The α_i and β_j terms correspond to "person" and "question" effects, respectively. The "Question" factor is a repeated measure. The $\alpha\beta_{ij}$ term of importance (their interaction). The error term, ε_{ijk}, results from the fact that the raters' assessments do not correspond perfectly. The summary table for this analysis follows:

Source	Sum of Squares	df	Mean Square	F-Ratio
Student	700.3	164	4.27	10.543**
Question	48.9	4	21.74	12.22**
Student x Question	368.6	656	0.56	1.39**
Residual Error	985.9	2434	0.41	
Total	2103.6	3458		

Eta(Question) = .152
**p<.01

3. Research Questions 11 and 12 can be examined with the same analytic procedure. The eleventh question concerns the difference (i.e., effect) between unassessed and assessed students. The twelfth question concerns whether or not these differences vary across institutions. A two-factor ANOVA design is called for. The Institution factor has eight levels; the pre-assessment versus post-assessment status factor is dichotomous. The first research question above corresponds to the pre-assessment/post-assessment main effect. The latter question corresponds to the institution-by-assessment status interaction. The ANOVA model for this analysis is:

$$X_{ijk} = \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk} ,$$

where X_{ijk}, the dependent variable, is the average (over the four raters) assessment given a student on overall self-awareness. The α_i effect represents differences between institutions. The β_j effects represent differences between assessment status. Their interaction corresponds to whether the difference between the two assessment states is the same across institutions. Error is composed of individual differences between students within an institution-assessment states cell. The cell means from this analysis are presented in Table B4-1. The summary table from the analysis follows:

Source	Sum of Squares	df	Mean Square	F-Ratio
Institution	99.2	7	14.2	3.11**
Assessment Status	58.5	1	58.5	12.84**
Interaction	58.1	7	8.3	1.82
Error	679.0	149	4.6	
Total	4425.4	164		

**p<.01

One should notice that due to radically different numbers of students within the cells of the ANOVA, the sums of squares of the effects and the interaction do not sum correctly. Whereas orthogonality of main effects and the interaction may be lost to some extent, the F-tests are reasonably accurate.

4. Research Questions 13 and 14 can be partly answered from one analysis. Both questions concern those students who have been assessed. Statistically, the first question looks at mean differences between the five questions. There are complications in the analysis, as noted in the text. The fourteenth question is concerned with a significant institution-by-question interaction. Therefore, the analysis called for is a two-way ANOVA, with institutions and questions as main factors. The factors have eight and five levels, respectively. The dependent variable is the average rating (of the faculty assessments) of an individual from a particular institution on a particular question. The ANOVA model follows:

$$X_{ijk} = \mu_{..} + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk} \quad .$$

In this case the α_i and β_j effects correspond to institutional and question main effects, respectively; $\alpha\beta_{ij}$ is their interaction. The error term, ε_{ijk} , is based on the pooled variance of individuals from a single institution on a single question. The means in Table B4-3 are the cell means from this analysis. The summary table follows:

Source	Sum of Squares	df	Mean Square	F-Ratio
Institutions	219.1	305	31.3	7.27**
Questions	59.5	7	14.9	3.46**
Interaction	39.1	28	1.4	0.32
Error	1141.2	265	4.3	
Total	5162.1	304		

**p<.01

As with the previous analysis, one should notice that due to unequal numbers of students in the cells of the analysis, the sums of squares do not total as one would expect, but the F-tests are reasonably accurate.

5. An additional set of B4 analyses was performed in order to test the Question x Treatment (assessment status) interaction. These analyses were two-factor analyses with the five questions as the repeated measures on the Student factor. Students, one factor, were nested within the other factor, Status of Assessment (Pre-assessment and Post-assessment). The ANOVA model for this analysis is:

$$X_{ijk} = \mu_{\cdots} + \alpha_i + \pi_{k(i)} + \beta_j + \alpha\beta_{ij} + \beta\pi_{jk(i)} + \varepsilon_{m(ijk)} ,$$

where X_{ijk} , the dependent variable, is the rating given a student who is either pre-assessment or post-assessment on one of the five questions by a faculty rater. The effect for Students, $\pi_{k(i)}$, indicates that students are nested under one of two assessment conditions (Pre- or Post-assessment). The latter Assessment Status factor is represented by the α_i effect. The

β_j effect corresponds to mean differences across the five questions. The error term is based on differences between the four raters who marked each student's free response answers. Since the assumption is made that the students are a random sample of the population of students, the Question-by-Students nested within Assessment Status Interaction is used as the error term in testing the Question-by-Assessment Interaction. The students nested within Assessment Status mean square serves as the error term against which the Assessment Status main effect is tested. Furthermore, the Question main effect is tested against the Question-by-Assessment Status Interaction. The summary table for this analysis follows:

Source	Sum of Squares	df	Mean Square	F-Ratio
Between Students	695.6	164	4.24	
Assessment Status	44.0	1	43.98	11.00**
Students/Assessment Status	651.6	163	4.00	
Within Students	420.7	660	0.64	
Questions	50.4	4	12.60	17.80**
Questions x Assessment Status	2.8	4	0.71	1.26
Question x Students/ Assessment Status	367.4	652	0.56	1.39**
Error	985.9	2434	0.41	
Total	2102.1	3258		

Eta(Assessment Status) = .151
Eta(Questions) = .155
$^{**}p<.01$

Appendix 5h: Study C1

1. This analysis was performed to calculate the average number of credits awarded in different institutional and occupational programs. These average values are presented in Table C1-2 and briefly discussed with Question 1. Additionally, information concerning the extent of institutional and occupational field differences was desired. Therefore, data from programs, one from each institution (Institution A, Law Enforcement; Institution B, Management, Data Processing, and Secretarial Science; Institution C, Law Enforcement and Data Processing; and Institution G, Law Enforcement and Secretarial Science) was analyzed in a two-factor ANOVA design, where the random-effect student factor was nested within the institutional program factor. The structural model for this ANOVA design is:

$$X_{ij} = \mu_{..} + \alpha_i + \beta_{j(i)} + \bar{\varepsilon}_{ij} \quad .$$

The dependent variable, X_{ij}, is total number of credits awarded each student. The first effect, α_i, represents the institutional program effect and the second, $\beta_{j(i)}$, represents the student nested within-institutional program effect. There were two independent faculty ratings of the number of credits appropriately awarded each student; this variation composed the within-cell or error term. It must be remembered that in this design, it is impossible to separate occupational field differences from institutional differences; the two effects are confounded as the α_i effect. The summary table for this analysis follows:

Source	Sum of Squares	df	Mean Square	F-Ratio
Between Subjects	1435.0	52	27.6	
Institutional Program	778.8	3	259.6	19.4^{**}
Student/Program	656.8	49	13.4	0.6
Within Subjects (Error)	703.3	29	24.3	
Total	2138.9	81	26.4	

Eta(Institutional Program) = .603

$^{**}p<.01$

2. The second set of C1 analyses refers to Question 2 and is illustrated in Table C1-3. For each institutional program, a two-factor fixed-effect ANOVA was performed. The first factor was the seven-level content subarea; the second factor was the Students factor. The structural model for this ANOVA design is:

$$X_{ijk} = \mu_{..} + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk} \quad .$$

The dependent variable, X_{ijk} , equals rating on the content subarea scale. There were two faculty ratings on each of the seven content subarea scales for each student; such faculty discrepancies comprise the error term. The α_i and β_j effects represent main factors as mentioned above. The primary purpose of this analysis was to investigate whether faculty raters assessed students differently by content area; this question is answered in the positive with a significant student by content subarea interaction. A secondary interest in the analysis involves whether average ratings on the content subarea scales differ. Since the results are quite inconsistent across institutions and occupational programs, two summary tables are presented. Neither should be viewed as representative.

School B, Management

Source	Sum of Squares	df	Mean Square	F-Ratio
Content Subarea	6.3	6	1.06	7.89**
Students	1.3	7	0.19	1.43
Content Subarea x Students	12.1	42	0.29	2.15**
Error	7.5	56	0.13	
Total	27.3	111		

Eta(Content Subarea) = .480

**p<.01

School G, Law Enforcement

Source	Sum of Squares	df	Mean Square	F-Ratio
Content Subarea	256.0	6	42.7	4.76**
Students	42.4	19	2.2	0.25
Content Subarea x Students	62.7	114	0.6	0.06
Error	1254.0	140	9.0	
Total	1615.1	279		

Eta(Content Subarea) = .398

**p<.01

3. The third set of C1 analyses refers to Question 3 and Table C1-4. The purpose of these analyses was to evaluate the extent of faculty agreement and consistency within each institutional program. The dependent variable in all analyses was total number of credits awarded. Two sets of analyses paralleled each other. The estimation of "interrater agreement" was

computed for every institution as one-way ANOVAs with students as the factor. The ANOVA model for each institutional program was:

$$X_{ij} = \mu. + \alpha_i + \varepsilon_{ij} \quad ,$$

where X_{ij} is the dependent variable, α_i represents the student main effect on true score, and error is composed of differences between the two faculty raters. It should be noted that in this analytic procedure, all mean differences between faculty raters are attributed to error. Therefore, a second analytic procedure was employed to discard mean differences between raters. In this second analysis, a two-way ANOVA design was used, with students as one factor and faculty raters as the other. The rater factor has two levels. The rater main effects could then be removed from the analysis. The resultant coefficients were labeled as coefficients of interrater consistency. The ANOVA model for this procedure was:

$$X_{ijk} = \mu.. + \alpha_i + \beta_j + \alpha\beta_j \quad .$$

The α_i effect represents student differences on true score. The β_j effect is the rater effect and is discarded from the analysis. The interaction term i serves as error. Below are the two parallel analyses for the Law Enforcement program at Institution A and the subsequent calculations:

Interrater Agreement

Source	Sum of Squares	df	Mean Square	F-Ratio
Students	1188.9	16	74.31	0.71
Error	1779.0	17	104.65	
Total	2967.9	33		

$$\text{Coefficient of Interrater Agreement} = 1 \frac{\text{MS(Error)}}{\text{MS(Students)}} = .00 \quad .$$

Interrater Consistency

Source	Sum of Squares	df	Mean Square	F-Ratio
Students	1209.0	16	75.6	4.9**
Raters	1502.2	1	1502.2	98.2**
Students x Raters	244.8	16	15.3	
Total	2956.0	33		

Eta(Rater) = .713

** p<.01

$$\text{Coefficient of Interrater Consistency} = 1 - \frac{\text{MS(Interaction)}}{\text{MS(Students)}} = .80 \quad .$$

4. The final C1 analysis is discussed under Question 4 and the results are presented in Table C1-5. The dependent variable is the difference between the number of credits awarded a student by one faculty rater and that awarded him by the other rater. Since each student was assessed by two raters, one difference appears for each student. The frequency distribution of these differences appears for each field and over all fields in Table C1-5.

Appendix 5i: Study C2

1. The average amounts of credit awarded to illustrative portfolios within institutions and occupational fields are discussed under Question 5 and are presented in Table C2-3. Due to the rather inconsequential amount of between-program variability, and the lack of comparability of cases across occupational fields, no analysis of variance was performed comparing institutional or occupational fields as factors. The numbers reported in Table C2-3 are simple averages, computed first over all raters for each illustrative case and then over all cases. The purpose of this procedure was simply to note if institutional differences were large. Institutional differences are discussed again in Table C2-4 under the fourth analysis.

2. A second analysis was discussed under Question 6 and presented in Table C2-5. For each of the five illustrative cases, the overall rating made by one faculty member was subtracted from that of the other faculty member from the same institution. These discrepancies were then summed over the five institutions within each content area. These frequency distributions appear in Table C2-5. A one-way ANOVA was performed comparing the three content areas. The summary table for this analysis follows:

Source	Sum of Squares	df	Mean Square	F-Ratio
Between Groups	142.58	2	71.3	15.09**
Within Groups (Error)	2466.51	522	4.7	
Total	2609.09	524		

**p<.01

Finally, the three distributions of differences are summed to form a total frequency distribution. This too appears in Table C2-5.

3. The third set of C2 analyses consisted of interrater reliabilities for each of the fifteen programs. These appear in response to Question 6 and in Table C2-4. As in C1, these were computed two different ways for each program. Interrater agreement coefficients consider between-rater differences as error. Coefficients of interrater consistency remove average rater differences from the analysis. For both analyses, the total number of credits awarded is the dependent variable. The ANOVA model for interrater agreement is:

$$X_{ij} = \mu. + \alpha_i + \varepsilon_{ij} \quad ;$$

the structural model for interrater consistency is:

$$X_{ijk} = \mu.. + \alpha_i + \beta_j + \varepsilon_{ijk} \quad .$$

In both instances, the effects for illustrative cases is α_i for the latter model; β_j represents rater variance which can be removed from the analysis. Conceptually, the between-rater differences, β_j, and error in the interrater consistency model form the error in the interrater agreement model. In the consistency model, error is the Case x Rater interaction. Below are illustrative ANOVA summary tables and the calculations of coefficients (taken from Institution E, Data Processing program).

Interrater Agreement

Source	Sum of Squares	df	Mean Square
Cases	342.67	4	85.67
Error	147.33	10	14.73
Total	490.00	14	

$$\text{Interrater Agreement Coefficient} = \frac{\text{MS(Cases)} - \text{MS(Error)}}{\text{MS(Cases)}} =$$

$$\frac{85.67 - 14.73}{85.67} = .83, .76 \text{ when stepped down.}$$

Interrater Consistency

Source	Sum of Squares	df	Mean Square
Cases	342.67	4	85.67
Rater	80.40	2	40.20
Error	66.93	8	8.37
Total	490.00	14	

Eta(Rater) = .405

$$\text{Interrater Consistency} = \frac{\text{MS(Cases)} - \text{MS(Error)}}{\text{MS(Cases)}} = \frac{85.67 - 8.37}{85.67} = .90,$$

.86 when stepped down to two raters.

4. Another set of C2 analyses were 15 two-way analyses of variance, one for each program. The purpose of the analysis, as explained after Question 7, was to determine whether faculty members were able to agree on assignment of credit to the seven content subareas within each field. Therefore, for each program, a two-way ANOVA was performed. The first factor, the case factor, used the five illustrative cases as levels within each occupational field. The second factor had seven content subareas as levels in each occupational field. There is some agreement for treating this factor as repeated measures, but the nature of the data suggests that the underlying assumptions probably fit the design used reasonably well.

The research question was answered positively (faculty did consistently place credits in content subareas) when a significant case-by-subarea interaction occurred. The "yes" response was placed in the appropriate cell in Table C2-6 when this happened. Likewise, a "no" was used when the interaction was not significant. The ANOVA model for this analysis was:

$$X_{ijk} = \mu_{..} + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk} ,$$

where X_{ijk} was number of credits assigned to a particular content subarea for an illustrative case. The ε_{ijk} term represents differences between raters within the same institution. The interaction term, $\alpha\beta_{ij}$, is the effect of most interest in this analysis. It represents the degree to which cases were successfully differentiated by content subarea. Below are two illustrative summary tables, one with a significant interaction and the other with a nonsignificant interaction.

Institution A, Management

Source	Sum of Squares	df	Mean Square	F-Ratio
Case	461.9	4	115.5	2.08
Subarea	1031.4	6	171.9	3.09**
Case x Subarea Interaction	351.0	24	14.6	0.26
Error	3889.3	70	55.6	
Total	5733.6	104		

Eta(Subarea) = .424

**p<.01

Institution B, Management

Source	Sum of Squares	df	Mean Square	F-Ratio
Case	2.07	4	.52	4.79**
Subarea	4.79	6	.80	7.39**
Case x Subarea Interaction	4.86	24	.20	1.88*
Error	3.78	35	.11	
Total	15.50	69		

Eta(Subarea) = .556
**p<.01
*p<.05

5. The bottom or overall row in Table C2-6 demanded an additional analysis for each of the three occupational subject matter areas. This analysis was a three factor, completely crossed, fixed effects ANOVA. The three factors were Institutions, Content Categories, and Cases. Hence, in this analysis there were eight possible variance components: the three main effects, three two-factor interactions, one three-factor interaction, and an error term. The ANOVA model for this analysis is:

$$X_{ijkl} = \mu_{...} + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk} + \varepsilon_{ijkl} \quad ,$$

where X_{ijkl}, the dependent variable, represents the credit rating made in a particular case in a particular content category by a rater from a particular school. The error term corresponds to differences between-raters from the same academic institution. The purpose of the three analyses is the same as the previous analyses, but here institutions are pooled within occupational areas. In each occupational area analysis, this ability to diagnose differentially is represented by the Content Subarea-by-Case Interaction. When this interaction is significant, the assessors are consistently agreeing with respect to the intracase differences, and a "yes" appears in Table C2-6. When the interaction is not significant, the assessors are not agreeing with each other in a statistically significant manner, and a "no" is recorded. For Data Processing, the interaction was significant; the interaction was not significant for either Management or Law Enforcement. Below are summary tables for the Data Processing and Management areas; the Management table is quite similar to that of Law Enforcement.

Data Processing Summary Table

Source	Sum of Squares	df	Mean Square	F-Ratio
School	31.6	4	7.91	3.00*
Subarea	188.8	6	31.46	11.95**
Case	202.9	4	50.73	19.26**
School x Subarea	74.4	24	3.10	1.18
School x Case	74.4	16	4.65	1.77*
Subarea x Case	287.2	24	11.97	4.54**
School x Subarea x Case	306.3	96	3.19	1.21
Error	921.8	350	2.63	
Total	2087.4	524		

Eta(School) = .123
Eta(Subarea) = .301
**p<.01
*p<.05

Management Summary Table

Source	Sum of Squares	df	Mean Square	F-Ratio
School	562.98	4	140.75	8.16**
Subarea	244.28	6	40.71	2.36*
Case	280.83	4	70.21	4.07**
School x Subareas	694.53	24	28.94	1.68*
School x Case	309.08	16	19.32	1.12
Subarea x Case	122.24	24	5.09	0.30
School x Subarea x Case	314.96	96	3.28	0.19
Error	4827.20	280	17.24	
Total	7356.08	454		

Eta(School) = .277
Eta(Subarea) = .182
**p<.01
*p<.05

Appendix 5j: Study C3

1. This study assumes the notion that similarity in the average of numbers of competences checked by a student and supervisor is a good indication of high consistency. The (dependent) variable in question is the number of checks made on task inventory checklists in three occupational fields. These analyses refer to Question 8. The number of faculty-student pairs is presented in Table C3-1, and the average number of checks by each are presented in Table C3-2. As can be seen in Table C3-2, two values were calculated for each institution and occupational field. The first was calculated based on checks made by students, the second used checks made by work supervisors. The averages presented in Figure VI-1 were first summed over institutions and then disaggregated into the six mutually exclusive content categories.

2. The number of tasks checked both by faculty members and their respective students are presented for each content category over the three occupational fields in Table C3-4 in response to Question 9. The percentages which are presented are calculated with the following formula:

$$\frac{\text{Number of Verified Checks}}{\text{Total Number of Student Checks}}$$

A verified check is one made by a student and also made by the faculty member in evaluating the student. Obviously, checks made by faculty and not by students are not used in this analysis. The purpose of the analysis is to show the degree of consistency between students and faculty checkings; this has particular implications toward the accuracy or validity of student self-assessment.

APPENDIX 6

Background Information on Institutional Studies of Expert Judgment

6a. Alverno College. Speaking Assessment - Instructions to Students

6b. Alverno College. Excerpt from Procedures Manual

6c. Alverno College. Excerpt from Assessment Manual

6d. Alverno College. Speaking Assessment - Individual Faculty Rating Form

6e. Alverno College. Speaking Assessment - Team Rating Form

6f. Alverno College. Speaking Assessment - Student Self-Rating Form

6g. Antioch College. English Skills Essay Grading Sheet

6h. Antioch College. Memo to Freshmen from Testing Office

6i. Antioch College. Scale for Writing Competence

6j. Antioch College. Memo to Faculty

6k. Antioch College. Writing Competency Questionnaire

6l. Thomas A. Edison College. The Pre-Graduation Conference: Information for Faculty Interviewers

6m. Thomas A. Edison College. The Pre-Graduation Conference: Information for Students

6n. Thomas A. Edison College. Application for Pre-Graduation Conference

C1L1: SPEAKING (FORM B_2) - Part I

(Recommended time - 3/4 hour)

Alverno College

Assessment Center

"Should college be open to anyone who wants to try it or should there be some criteria operative to screen students before they enter college?" Take a position on either side of this question and prepare a 3-5 minute speech to convince an audience of the worth of your position.

Your audience will be a group of high school students who are looking into their own futures. Use specific instances from experience and observation to support your ideas and clarify them for this audience.

Spend about 5 minutes going over the above paragraph so that you have a clear picture of the kind of speech you are asked to give.

Take about 15 more minutes to prepare your speech. Then present it to the audience provided, as if they were a high school assembly. Although you may use notes, your task is to give a speech, not read a paper. A videotape operator will tape your presentation so that you can see it to judge your own strengths and weaknesses.

Excerpt From Procedures Manual

Alverno College

Assessment Center

Directions for Administering C1L1 Instruments

GENERAL INTRODUCTION FOR ALL C1L1 INSTRUMENTS WHEN ADMINISTERED TO GROUP PRIOR TO ENTRANCE (Some standardized instruments may also be administered at this time. Note to administrator: Some students may have already taken these instruments before they entered Alverno. CHECK THIS OUT AHEAD OF TIME. DO NOT HAVE THEM REPEAT IT.)

Ordinarily C1L1 instruments are administered to incoming students in a group. When it is necessary to administer them individually, no general introduction is necessary.

Tell participants that:

> (At Alverno learning is based on development of competence. We believe that a person can develop competence only if she herself understands what strengths she is building on and what weaknesses she needs to eliminate. During all of your college career, we will assist you as students to keep refining your ability to read, write, speak, listen, and use statistical information as an educated adult. Therefore from the start we work to identify your strengths and weaknesses and assist you in identifying them yourself.
>
> We will begin that process today. The results will help us clarify some of the necessary emphasis in your academic program and study time next year. Today we are arranging some time to help you as entering college students find out exactly what you can do in the college skills of analytic reading, analytic listening, speaking, writing, and graphing. For example, today among other things, you'll be writing a letter and giving a short speech. If it's clear from your speech and your letter that your speaking and/or writing are sufficiently advanced, you will be exempt from the one-hour Writing and Speaking class that is required of freshmen. If it's clear that, in addition to freshman Writing or Speaking class, you need to spend some extra time to brush up, you will be directed to do so.
>
> Your assessment experiences today will differ in several ways from tests you're used to taking:
>
> 1) They do not depend on memory.
> 2) They do not test the extent of your knowledge.
> 3) They do not involve failure. If you can really do a better job than your performance on them suggests, you can try again later.

They will merely evaluate what you can *do* in analytic reading, listening, writing, speaking, and graphing so that we can figure out what college skills you already have a good start on.

How will that work?

You will be asked to:

A. Read a two-page article and answer questions on the article--with the article still in your possession

B. Listen to a speech given on a videotape, and answer questions on the speech--from notes which you took as you listened

C. Prepare and give a short speech on an assigned topic that does not require special knowledge

D. Write a letter on an assigned topic related to your past experience

E. Read two to three graphs, answer questions on the graphs, and construct a simple graph

F. Finally, because we want you to understand how well you do so that you know what you have to learn in order to improve, we will ask you to look at your own work in each of the above cases except your speech, and judge how well you did. We'll take some time in September for you to take a look at your speech and evaluate it.)

Note: There should be a time lapse between Part I and Part II of these instruments so that a student has an opportunity to get at least the *start* of some distance from her own work. If they are given in a group, all the Part II's should be last, in the following order: Graphing, Reading, Listening, Writing. (For Speaking, Part II is given in the Feedback session.) If the C1L1 instruments are administered individually, the student should make 2 separate appointments--one for Part I and one for Part II--in Speaking and Writing. The others (Reading, Listening, and Graphing) need not have a time lapse.

Alverno College
Assessment Center

Directions to the Administrator for C1L1 Speaking - Part I

C1L1 Speaking, Part I

TIME ALLOTMENT: Preparation time is detailed on student direction sheet--the instrument itself; taping time is hard to determine; in general, allow eight students for a one hour taping session.

MATERIALS NEEDED: Part I of the assessment instrument
scrap paper--pencils
videotaping equipment and operator
(Secure equipment and operator through the Media Center. Contact them far in advance of need with date, time, and kind of equipment requested)

1. Tell the participants:

(You will now be asked to prepare and give a short speech on an assigned topic that does not need special knowledge. Prepare notes to speak from; but do NOT write out a paper to read. Do not worry about the exact words you will say, but concentrate on getting something across to the audience to whom you will be talking.

In order for both you and us to get a clear idea of how well you do, we need to videotape you. Therefore you will give your speech in the media studio.)

Alverno College: Excerpt from Assessment Manual

SPEAKING ASPECT OF COMPETENCE 1, LEVEL 1 (C1 L1)

The speaking aspect of C1 L1 asks that a student:

> Identify own strengths and weaknesses as initiator in an oral communication situation involving an audience which, with situations in the other aspects of C1 L1, can be called varied.

For a student to do this she must:

> As an initiator make an oral presentation in which she uses appropriate voice techniques and reinforce the message by general bodily response
>
> Select ideas, organize them in relation to a focal idea, and express them so that in a communicable form they contribute to an identified purpose--for an identified audience
>
> Develops ideas with sufficient appropriate examples to clarify and sufficient appropriate evidence to substantiate her message

To demonstrate these behaviors, students give a short speech for which they have been given some preparation time and directions as to situation, purpose, and audience (See Appendix B, p. 4). A videotape is made of their performance. The judgments that we assessors will be asked to make will be solely on the basis of that tape. As we view it, we will try to determine whether certain essentials of good speaking are strengths or weaknesses in a given student's performance. The skills we will watch for are six, which are further specifications of the three behaviors listed above (See Appendix B, p. 5). Because of the importance of delivery in speaking, #1-4 are subdivisions of the first behavior above (voice techniques and non-verbal reinforcement):

Item #1: Clear ARTICULATION

> Does the student speak distinctly, without joining sounds that should be separated, without adding extraneous sounds or omitting necessary ones, and without distracting by any other distortion of sound?

The most common signal of unsuccessful articulation is what people often refer to as "muttering." This might mean that the speaker is slurring consonants like p, b, t, d, k, g or clipping vowels like long a's and o's. In a given speech, do we as audience hear individual words easily? If so, the speaker has clear articulation. Of do we lose a word every here or there--or in an extreme case throughout--without special effort? If so, articulation for that speaker is a weakness.

A certain sign that any voice technique is successful is that the hearer is not aware of it because it contributes effectively as a means to an end--a well delivered total message. On the other hand, if we are not accustomed to hearing good speakers we might momentarily be distracted by some aspect of technique that is especially well handled. If we gradually hear it as part of the total experience, it is probably integrated into it. With articulation, however, if it calls attention to itself and its clarity, it may be that it is over precise. In that case it is a weakness in the opposite direction. For example if a speaker calls attention to letters like final *t*'s of words like not and write, or middle *t*'s in words like letter, pronouncing them all like the initial *t* of tack and tomato, that speaker's articulation is one of his or her speaking weaknesses.

Item #2: Sufficient VOICE PROJECTION

Does the individual use his or her speech mechanism with enough forcefullness to suggest care about holding the attention of the audience?

In judging voice projection as a strength or weakness of a given student, we will be looking at the effect, in her delivery, of what might be a variety of misuses of her speech mechanism. Nevertheless it is not our task to analyze the causes. We make a yes-no evaluation of the effect.

Does the student project her voice? That is, does she, in her production of sound, seem to send the message forth rather than hold it to herself? Does she use her speech mechanism to create resonance? Do her lips, by their vigorous action, facilitate rather than obstruct the carrying power of her voice? If our answer to all of these is yes, then voice projection is a strength in her performance. If we can say no to any one of them, it is a weakness.

Item #3: General BODILY RESPONSE that reinforces message and tone (includes use of such things as eye contact, facial expression, and movement)

Does the student reinforce and/or vitalize her message with appropriate non-verbal actions?

No matter how effective a speaker's use of his or her voice might be, if the other physical aspects of that person's bearing do not reinforce it and appear as part of a total integrated experience, their failure diminishes and obstructs the power of the voice. Therefore, in assessing a student's reinforcement of message and tone by bodily response we need to examine the contribution of specific elements to the total communication.

Because, in the situation we will be assessing, students will be standing behind a lectern, their bodily movement will be limited. What elements, therefore do we judge? First of all, eye contact. Are the speaker's eyes more in contact with her audience than with anything else that causes her to look down rather than out (including her paper, if she uses one)? Next, posture. Does the student stand in a way that suggests she has something worthwhile to say? Or does she slouch or shrink as if apologizing for speaking, or at least as if she does not expect an audience to listen?

If she is using gestures, do they emphasize her thoughts and feelings? Or do they detract by closing off voice power (hand over mouth) or facial expression (hand over other parts of face)?

Finally, facial expression. Does it vary? Does it help bring the speaker's words to life? Or does it call attention to itself by wondering whether it is going to change or come to life?

In order to say that a student's general bodily response is a strength of hers in reinforcing message and tone, one would need to be able to answer yes in all of these areas. An exception of course is the area of gestures. If she uses them, we must judge them. If she does not, we ignore the question.

Item #4: INTEREST in message expressed by VARIATIONS in PITCH, VOLUME, and PAUSE

Does the student's use of pitch range, of pause, and of volume range contribute to an ability to hold her audience?

A speaker may cause a feeling of discomfort or strain in the audience by his or her failure to vary pitch or by her tendency to vary it excessively and without relation to meaning of message. He or she may also cause such a feeling by choppy rhythm created by pause that is too regular or too frequent. Finally, the speaker may misuse volume by not speaking loudly enough and thus causing strain, or by speaking too loudly and thus causing discomfort. All of these situations would immediately identify weakness in the skill we are considering here.

However, even if the speaker causes no discomfort except that of monotony by failing to vary pitch, volume, and pause meaningfully, he or she is weak in this specified area of creating interest. If, with a clear absence of the obstacles enumerated, the student speaker indeed captures the interest of the audience by her use of pitch, volume, and pause, we can count as a strength the item we are evaluating here.

Item #5: ORGANIZATION of total message through FOCAL IDEA and CLEAR LEADING of audience FROM PART TO PART

Does the student make the audience aware throughout of why she is saying all the things she is--in terms of clarifying a central idea--and of how she is relating them to that idea and to each other?

As we listen to a speaker, we should be aware of the main point he or she is trying to make. Or at least have a suspense that causes us to stay with the speaker until we reach that awareness. We should realize how the statement the speaker is making at each given moment is related to that idea. If the speech we are hearing does not enable us to do that, the speaker is not demonstrating the skill we are concerned with here.

Further, that speaker should be giving us clues as to the direction, and--at a given moment--the position, within that direction, of whatever thought he or she is stating or developing. There are a variety of techniques available that enable a speaker to signal direction and position: Therefore . . . on the other hand . . . however . . . nevertheless . . . although . . . first, second, third . . . consequently . . . but . . . not only/but also . . . yet . . . and many others. However important these might be in movement from sentence to sentence, however, their individual use is not enough unless they are a clear part of a larger pattern like one of time, significance, suspense, space, or logic. Do we sense such a pattern and do we find in the words of the speaker, the map through it? If so, this skill is clearly a strength.

Item #6: APPROPRIATE USE of EXAMPLES to illustrate idea and convince of its worth

From our experience with recognition of example in the listening competence, we no doubt remember that an example concretely represents some general concept or situation or type or operation to which it is related.

Does the speaker, we then ask, use examples? If so, do they clarify the idea by bringing it into the realm of concrete experience for the audience and/or do they serve to make worth mentioning, the abstract point they are exemplifying?

In the student speeches that we will be assessing the speaker's task is to nominate someone who has taught her, as Outstanding Teacher of the Year. It will probably not be difficult to listen for examples, for they will surface, if they are there, without much analysis on our part. However, the mere presence of examples is not enough. Do they really clarify the idea they are used to exemplify? Speakers will probably be using examples to clarify the qualities they attribute to the teacher they nominate. Do the examples they use help us understand exactly in what the teacher has qualities they describe? Do the examples move us to agree that this teacher is worth nominating? A yes answer to these questions chalks up this skill as a strength for the speaker we are assessing; a no answer to them chalks it up as a weakness.

The questions proposed above for all of the six indicated items are the questions we need to ask ourselves as assessors when we view each student's videotape. We judge her performance in terms of each of the six identified skills as strengths or weaknesses, without necessarily making a general observation on the speech as a whole. The level of correspondence between her judgment and our common judgment will automatically indicate whether or not she has achieved Level 1, which is the ability to identify her own strengths and weaknesses.

Editor's Note: The appendices referred to in this Assessment Manual excerpt are not included in The CAEL Validation Report.

Alverno College
Assessment Center

6-9

Assessment Record
Individual

C1L1: SPEAKING

Below is a list of some essential characteristics of good speaking. Look at and listen carefully to the video tape of the speech. Then check (✓) each one yes or no --depending on whether you judge that the student has adequately done what is described. Indicate evidence wherever you can.

1. yes ___ no ___ Clear Articulation: Did she speak distinctly, without habitually joining sounds that should be separated, without adding extraneous sounds or omitting necessary ones, and without distracting by any other distortion of sound?

2. yes ___ no ___ Sufficient voice projection: Did she use her speech mechanism with enough forcefulness to suggest care about holding the attention of the audience and to make the audience feel that she was talking to them?

3. yes ___ no ___ General bodily response that reinforces message and tone (includes use of such things as eye contact, facial expression, and movement): Did she reinforce and/or vitalize her message with any appropriate non-verbal actions? (Judge contact and facial expression only, unless she happened to use any other gestures.)

4. yes ___ no ___ Interest in message expressed and evoked by variations in pitch, intonation, volume, and pause: Did her use of pitch range, of pause, and of volume range help avoid monotony by creating clear contrast and emphasis on some ideas?

5. yes ___ no ___ Organization of total message through focal idea and clear leading of audience from part to part: Did she make it clear (for the sake of her audience) throughout why she was saying all the things she was -- and of how she was relating them to the focal idea and to each other?

6. yes ___ no ___ Appropriate use of examples to (a) illustrate ideas and (b) convince of the worth of ideas: (a) Did she use examples that clarify a given idea by relating it to concrete experience familiar to the audience? (b) Did she use examples that show how her abstract ideas are worth mentioning?

Name ____________________

Form ____________________

Date ____________________

Assessor ________________

0675

Alverno College
Assessment Center

6-10

Assessment Record
Team

Cl Ll - SPEAKING (AC)

Directions: Check yes or no for each item. If you check no, indicate evidence to support your decision.

	Yes	No	Student Evaluation	Correlation	Evidence
1.					
2.					
3.					
4.					
5.					
6.					

To Be Filled in at Assessment Center

Credentialized ____________________

Not Credentialized__________________

Date____________________________

Feedback By_______________________

Name____________________

Date____________________

Assessor_________________

Date____________________

0675

Alverno College Assessment Center

C1L1: SPEAKING (FORM B_2) - Part II

Below is a list of some essential characteristics of good speaking. Look at and listen carefully to the video tape of the speech you gave. Then check (✓) each one yes or no--depending on whether you judge that you've adequately done what is described. Indicate evidence wherever you can so that your understanding of the situation is clear when it is considered for credentialization.

1. yes___ no___ Clear articulation: Did you speak distinctly, without habitually joining sounds that should be separated, without adding extraneous sounds or omitting necessary ones, and without distracting by any other distortion of sound?

2. yes___ no___ Sufficient voice projection: Did you use your speech mechanism with enough forcefulness to suggest care about holding the attention of the audience and to make the audience feel that you were talking to them?

3. yes___ no___ General bodily response that reinforces message and tone (includes use of such things as eye contact, facial expression and movement): Did you reinforce and/or vitalize your message with any appropriate non-verbal actions? (Judge contact and facial expression only, unless you happened to use any other gestures.)

4. yes___ no___ Interest in message expressed and evoked by variation in pitch, intonation, volume, and pause: Did your use of pitch range, of pause, and of volume range help avoid monotony by creating clear contrast and emphasis on some ideas?

5. yes___ no___ Organization of total message through focal idea and clear leading of audience from part to part: Did you make it clear (for the sake of your audience) throughout why you saying all the things you were--and of how you were relating them to the focal idea and to each other?

6. yes___ no___ Appropriate use of examples to (a) illustrate ideas and (b) convince of the worth of ideas: (a) Do you use examples that clarify a given idea by relating it to concrete experience familiar to the audience? (b) Do you use examples that show how your abstract ideas are worth mentioning?

Name_______________________________

Antioch College

GRADING SHEET

ENGLISH SKILLS ESSAY

Student's number ____________ Grader's number ____________

Scoring criteria:

1. Correctness in spelling, punctuation, grammar and format....................
2. Accuracy in the choice of words......
3. As much clarity in every statement as the matter permits..................
4. Reasonable order in the construction and arrangement of paragraphs........
5. Reasoning, use of logic or evidence to support conclusions..................
6. Pointedness or relevance to the topic.

	1 low	average	3 high
1.			
2.			
3.			
4.			
5.			
6.			

Instructions: For each of the above scales, place a check mark in the box which best represents this paper.

Over-all Rating

LOW AVERAGE

|____|____|____|____|____|____|____|____|____|____|____|____|____|____|

Instructions: Place a check on the line to represent how you feel this paper compares with the other papers in the group you are grading. Use the whole scale.

Comments: Please make any comments about the paper which you feel will be helpful to the student or the tutor.

Antioch College

July 1967

To: Freshmen

From: Testing Office

Re: Examination and Improvement of Student Writing

The following policy on examining and improving the quality of students' written work was approved by the General Education Committee, Tuesday, June 4, 1963, and by the Educational Policy Committee on September 3, 1963.

Standards of Prose

Every student admitted to Antioch College is expected to be able to communicate ideas accurately, clearly, coherently, interestingly, in writing. The specific aim of the English Skills Examination, the Continual Check on Writing, and the Tutorial Service is to ensure that students shall have mastered that art, the ability to write formal expository prose.

Such prose, dignified without being pompous, is the kind of writing one finds in the better scholarly journals and in a good paper for academic work. A formal essay should have the following qualities:

1. Correctness in spelling, punctuation, grammar and format
2. Accuracy in the choice of words
3. As much clarity in every statement as the matter permits
4. Reasonable order in the construction and arrangement of paragraphs
5. Independent thinking
6. Pointedness or relevance to the topic

The diction of formal expository prose is usually void of such things as contradictions, second-person pronouns, boring repetitions, slang, cliches, triteness, and colloquialisms.

Once a student has mastered the art of writing in the formal style of civilized discourse, he may carefully depart from the norm in various ways in order to achieve any needed rhetorical stress which conventional diction cannot bear. And he should then experiment in other kinds of writing, as inclined. But in the interest of conserving the power of language as an instrument of thought, a means of communication and a civilizing agent, we encourage people at first to imitate rather the style of Walter Lippmann than that of Ernest Hemingway; we cherish more the skill to write an exact parody in the manner of Jonathan Swift's A Modest Proposal for Preventing the Children of Ireland from Being a Burden to Their Parents or Country than the knack of imitating Holden Caulfield's word-flow in The Catcher in the Rye.

Every ordinary paper, whether an essay for academic work or a cooperative job report, should meet these standards. If it does not, the instructor may refuse to read it, or he may require that it be rewritten for appropriate credit. And at any time, if the instructor or advisor ascertains that a student needs help in writing, he may either urge or require the student to consult the tutor.

Antioch College

Office of Educational Evaluation
and Research
Form B

ESSAY SHEET

STUDENT ______________________ READER ______________________

Scale for Writing Competence

Clarity, vocabulary, grammar		Organization and expression of ideas; style
Writing is clear and precise with a broad vocabulary appropriately used, and excellent grammatical usage.	9	Writing is well organized. Appropriate language is used to convey ideas. Style shows variety and imagination.
	8	
Clarity is adequate, vocabulary good, and grammar correct.	7	Writing is organized and ideas are clearly expressed. There are some indications of personal style.
	6	
Clarity of writing is generally adequate, vocabulary is somewhat limited, and minor errors in grammar and spelling are sometimes made.	5	Writing is organized, with most ideas clearly expressed, but on occasion the organization is not completely coherent and the language unconvincing. Style tends to be pedistrian.
	4	
Writing is sometimes unclear with words at times used inappropriately and misspelled. Major grammatical errors are occasionally made.	3	Writing is hard to follow. Positions on issues are unclear because of poor expression of ideas. (Or problems of organization and expression are avoided through oversimplication.) Paper lacks style.
	2	
Writing is often incoherent and disjointed. Vocabulary is extremely limited, words are misused and spelling is poor; major grammatical errors are often made.	1	Writing leaves the reader confused. Organization is lacking and language fails to communicate ideas.

TO READERS: Above is a nine-point scale to use in judging the competence in writing demonstrated by the student in his essay. To the left, the scale has been defined in terms of clarity, vocabulary, and grammar; to the right, in terms of organization, expression of ideas, and style. You are to make one overall judgment. Circle the number which best expresses your overall judgment. Use the back of this sheet to make comments explaining your rating, useful to the student and his advisers. It is important to make these comments.

Antioch College

Office of Educational
Evaluation and Research
October 31, 1975

To: Yellow Springs Faculty

From: Peggy Brown

Our office is currently conducting a study of the use of expert judgment in the evaluation of writing competency, sponsored by a grant from the Cooperative Assessment of Experiential Learning project. Part of the study involves checking professors' assessment of writing skills for fall entrants in their courses.

As a preliminary step we need to know whether you have fall entrants in your courses and how much writing is involved in each course. Will you please fill out the attached form and return it to our office by Friday, November 7?

Thank you.

- -

Name:

Course #	Much Writing	Some Writing	No Writing
________	_____	_____	_____
________	_____	_____	_____
________	_____	_____	_____
________	_____	_____	_____

_____ I do not have any courses in which fall entrants are enrolled.

_____ I do not have any courses which involve writing.

Antioch College

Office of Educational
Evaluation and Research
Fall 1975

WRITING COMPETENCY QUESTIONNAIRE

Student:

Adviser:

Did you discuss the evaluation of his/her writing with this student?

_____ Yes _____ No

Comment: If yes, what suggestions were made for improving writing skills?

Has the student followed through on suggestions?

_____ Yes _____ No

Comment:

Have you been able to identify any improvement in writing skills?

_____ Yes _____ No

Comment:

Thomas A. Edison College

The Pre-Graduation Conference: Information for Faculty Interviewers

BACKGROUND INFORMATION

The Pre-Graduation Conference is a face-to-face discussion, lasting approximately one hour, between each student candidate for the B.A. degree and two college faculty members. One faculty member is selected to discuss with the student certain topics in his or her "area of concentration" (Humanities, Social Sciences, or Natural Sciences). The second faculty member covers more briefly topics in one of the other two areas.

The essential purpose of the conference is to ensure that every degree candidate demonstrates an acceptable grasp of basic concepts both in the Area Concentration and in the second area he or she selects.

THE RESOURCES AVAILABLE TO THE FACULTY INTERVIEWER

As a faculty interviewer, before the conference you will be sent a copy of the student's application, which will give the student's suggestions for topics to be covered and will indicate something of the student's background in these topics. You may, at your discretion, go beyond the topics listed, in exploring the student's grasp of basic concepts.

You will also be sent a copy of the evaluation form listing the courses for which the student has earned credit, and you will be given further information if the student has opted for a specialization (See pp. 24-25 of the 1975-1976 Thomas A. Edison College Catalog).

You will also be given a copy of The Pre-Graduation Conference: Information for Students, to enable you to see the conference from the student's point of view.

In addition, the College is in the process of developing a file of general questions that the faculty interviewers may ask the student. This file is intended:

1. to indicate to the faculty interviewer the level of generality expected by the College in the questions asked

2. to indicate to students, by making public a sample of questions from the file, what they may expect at the Pre-Graduation Conference

3. to indicate to Counselors at the College, and to interested educators outside of the College, the "flavor" of the Pre-Graduation Conference process

4. to provide a means of standardizing the Conference process

It is intended that faculty interviewers will be able to choose among questions in the file to ask the student, but not be limited to using only such questions.

The College would be happy to receive from faculty interviewers suggestions for additional questions to be included in the file.

THE EVALUATION

As an outcome of the conference, each faculty member is to arrive at a judgment of "pass" or "defer with recommendations" in the area for which he or she has been given responsibility. The faculty member may, if he or she wishes, consult with the other faculty member and with the staff member present before making this judgment.

"Pass" means the faculty member believes the student has an acceptable grasp of basic concepts in the area under consideration.

"Defer with recommendations" means the faculty member does not believe the student has an acceptable grasp of basic concepts. This judgment, when given, is to be accompanied by specific suggestions for further reading, courses, etc., intended to bring the student to the "pass" level.

In addition to recording his or her judgment on the appropriate form and signing that form, each faculty member is to briefly record on a second form some of the considerations that led to this judgment. Since this second form is now in an experimental phase, the directions for its use will not be given here.

ORGANIZATIONAL DETAILS OF THE CONFERENCE

To avoid scheduling problems when a number of Pre-Graduation Conferences are given on the same day, the entire conference process for any given student should be completed within a one-hour period. The suggested timing for each component is as follows:

5 min. - Interviewers and College staff review the student's Application for Pre-Graduation Conference and other background materials and plan a general strategy for the conference (e.g., selection of questions to be asked, suggested "lead-ins" for discussion).

3-5 min. - Student introduced to faculty and staff participants; informal social conversation to put student at ease.

25 min. - Conversational questions and discussion in student's area of concentration, led primarily by faculty member in that area.

15 min. - Questions and discussion of topics in secondary area, led by other faculty member.

5 min. - Student excused from room; interviewers and staff discuss student's performance and arrive at "pass" or "defer with recommendations" decisions.

5 min. - Student returns and is informed of conference outcome. This "debriefing" period can include commendation for any areas of exceptional performance, as well as suggestions for additional work or study in areas where some deficiency was noted.

Naturally these suggested timings cannot always be adhered to rigorously. However, to the extent possible, each interview should be conducted within these general time allotments. The staff member present will help guide the faculty interviewers in this respect.

All conferences are cassette-recorded for inclusion in the files of the Office of Test Development and Assessment and for internal study to improve the conference process, except where the student prefers not to have the conference recorded. The tape recorder will be operated by the College staff member present.

SUGGESTIONS TO THE INTERVIEWER

The following suggestions represent a distillation of our experience in providing an appropriate psychological atmosphere and insuring that the discussions focus on the student's knowledge of basic concepts. These suggestions--or any other listing of "do's" and "don'ts"--cannot, of course, fully express all of the elements that make for an effective highly satisfactory interview.

1) Every effort should be made within the first few minutes of the interview period to put the student at ease. A smile, friendly greeting, handshake, and two or three minutes of "small talk" are very important for this purpose. Even students who outwardly appear at ease may be quite tense. Do not let the discussion jump immediately into areas of substantive content. A brief "social conversation" period is encouraged as a matter of course with all candidates; students who show visible anxiety may require a somewhat longer period.

2) It is generally not helpful for the interviewer to verbalize that "this is just a friendly conversation," "there is no need to be anxious about the interview," "we want you to put yourself at ease and not worry about the questions we are going to ask," etc. These and similar instructions to the student about "relaxing" "enjoying the experience," and so forth may be counterproductive and generate more anxiety than the student had initially. The much preferred approach is for the interviewer to show indirectly--through a relaxed manner, friendly small-talk, and so forth--that the conference situation will not be a threatening one.

3) Wherever possible, the substantive portion of the interview should flow naturally and without an obvious break from the preceding "small-talk." A useful bridge between the two is to ask the student about his or her employment, recent course work, or current reading in a given area. Any of these can provide a useful lead-in to more substantive questions. But be careful not to prolong this transitional period unduly by becoming so engrossed in details of the student's job, reading interests, etc. that you neglect questions relating specifically to subject matter areas.

4) Since the major purpose of the interview is to have the student demonstrate knowledge of basic concepts in the areas under discussion, avoid the quite natural tendency to present your own views on a given subject, to "teach" the student in the course of the interview, or, in some cases, to discuss with the other faculty member a particular concept or theoretical point. Once the introductory period of social conversation has passed, the student should "have the floor" for the most part, and the interviewers should talk only enough to give the student appropriate lead-ins to the concepts or areas they would like talked about, or to redirect the conversation as necessary in the course of the interview.

5) Much of the interview will inevitably take the form of questions by the interviewers, with responses by the student. However, for variety, and to reduce as much as possible the "interrogation" aspects of the interview, you might, wherever possible, use *comments* rather than questions to elicit student responses. For example, a comment such as "I imagine you can see some relationships between Skinner's theory of reinforcement and your own role as supervisor" has the same effect as, but is much less blatant than, "What is the relationship between Skinner's theory of reinforcement and your work as supervisor?"

6) Avoid comments which have an *evaluative* implication, such as "that's a good answer" or "correct." Expressions such as "that's an interesting (note: not *good*) idea" can be used instead to encourage the student.

7) In general, the student should be addressed by only one interviewer at a time, while the second interviewer listens. However, it may occasionally be appropriate for one interviewer to ask a question or comment during the "official time" of the other. This occasional interchange makes for more natural communication. But use it sparingly, so that the interview does not become a sort of verbal "ping-pong match." If you interrupt a two-way conversation, make sure that your question is germane and be careful that you are not interrupting a chain of questions the first interviewer is deliberately exploring.

8) Avoid long pauses or embarrassing confusion on the student's part. If the student is at a complete loss for words, or is obviously having difficulty in responding adequately in a particular area, you should quickly and as unobtrusively as possible refocus the question or change to another subject. A few words to the effect of "no problem, let's go on to (subject)..." will help to smooth this transition.

9) It is generally not a good idea to take frequent written notes during the interview. However, you may want to jot down one or two key words (in an unobtrusive way) to help you to summarize the student's performance at the end of the conference and to make recommendations for further study.

10) Toward the end of the interview period, spend a few minutes "winding down" the discussion to avoid an abrupt and possibly disconcerting termination. One technique is to say that only a few minutes remain, during which the student may wish to discuss some additional area of his or her own choosing. This will give a feeling of completeness and adequate-opportunity-to-talk on the student's part.

Thomas A. Edison College

The Pre-Graduation Conference: Information for Students

What is the Pre-Graduation Conference?

The Pre-Graduation Conference is a face-to-face discussion between the student and two faculty members. It is required of all candidates for the B.A. degree.

The purpose of the conference is to ensure that you have an acceptable grasp of basic concepts both in your chosen Area Concentration (Humanities, Social Sciences or Natural Sciences/Mathematics, including your specialization if you have one), and in a second area of your choice (Humanities, Social Sciences, or Natural Sciences/Mathematics).

How is the conference scheduled?

Soon after you have accumulated 80 semester hours in the B.A. program, the Director of Liberal Arts Programs will review your current evaluation form. If a reasonable portion of your Area Concentration and Liberal Arts electives has been completed, you will be sent an application form for the Pre-Graduation Conference. You should fill it out and return it to the Edison College Office of Test Development and Assessment. The conference itself can be planned for several weeks or months later, depending on individual circumstances. But once the form has been sent to you, you should begin to fill it out as soon as possible.

On the application form indicate those days of the week and times of day you are NOT available for the conference. The College will make every effort to arrange a convenient date and time for the conference, but there may be difficulties involved in bringing together you, the faculty members, and a College staff member. If the conference must be held on a day you work, the College will be glad to explain the importance of the conference to your employer.

How should I prepare for the conference?

In the conference you will be discussing some of the major concepts, issues, debates, theories or problems within your area of concentration, and, to a lesser extent, within the second (non-concentration) area you select.

The conference application form asks you first to check off your Area Concentration and the secondary area you select for the conference. Then you are to list several topics within each of these two areas in which you have a particular interest and background. The next page gives some sample topics as suggestions. You are free, of course, to list other topics than these.

HUMANITIES

Religious basis for the Reformation
Journalism: newswriting, copy editing, feature writing
Aristotle's world view
Comparison between Medieval and Renaissance art
Music as a communicative art
Vonnegut's social criticism
Spanish culture in the New World
Ionesco and the Theatre of the Absurd

SOCIAL SCIENCES

Ethnicity - race relations - minority groups
State and local government
Population ecology
Freudian vs. behavioral theories
The rise of Fascism
Technology and its effects on society
Social psychology of organizations
The growth of civil liberties within the framework of American History

NATURAL SCIENCES/MATHEMATICS

Basic principles of aerodynamics
Implications of the theory of evolution
Practical applications of calculus
Oceanography - oceans as a potential source of food, fuel, power
Alternate number systems
Mathematics in Western Civilization
Human nutrition
Astronomy - The Copernican Revolution

It is important for you to give thought to the topics you list. The College will use the information you provide to select appropriate faculty members for your conference. The faculty members in turn will use this information in planning and conducting the conference (although they need not limit themselves just to the topics you list).

You are asked to give the names of at least two books or articles you are familiar with for each topic you list. The faculty members may also find useful information concerning how your work, study, or leisure activities relate to each topic.

To prepare for the conference you should:

1) carefully prepare the application, with the purpose of the conference in mind.

2) review the application, if you wish, with the College counseling staff.

3) refresh your memory of courses taken or other learning experiences, to recall major concepts and principles, as opposed to minute, specific details.

How is the conference conducted?

Almost all Pre-Graduation Conferences are held at the College. Plan to arrive there at least fifteen minutes before the scheduled meeting time. A College staff member will be on hand to greet you and introduce you to the other conference participants. The staff member may also be present at the conference, along with the two faculty members.

The conference will be run on a conversational basis. You will not be required or expected to make a prepared speech. About two-thirds of the discussion will be addressed to topics in your area of concentration, and one-third to topics in the other selected area. At the end of the discussion period, the faculty members will review the conference with each other, and will then discuss with you the overall outcome. They may also have suggestions to help you plan the remaining learning activities in your degree program.

If serious deficiencies in your overall preparation become evident at the conference, the College may require additional work prior to graduation. This has happened in less than 10% of the conferences held so far. If you have seriously prepared for the Pre-Graduation Conference along the lines suggested, you should anticipate a positive recommendation for you to continue and complete the degree program.

The College would like to make a cassette recording of each conference. If you feel that the presence of a small cassette recorder would adversely affect your performance in the discussion, this part of the process can be omitted. However, we hope that a majority of students will allow us to gather this kind of record for research and program improvement purposes.

Interview recordings and conference notes are kept in the Office of Test Development and Assessment, and are not made available outside of the College. A notation is made on the student transcript when the Pre-Graduation Conference requirement has been successfully completed.

What fees are charged for the conference?

Any student enrolled in the College's B.A. program may participate in the Pre-Graduation Conference at no additional fee, when the conference is held at the College. Students who cannot attend a conference at this location may request special arrangements for another site. At its discretion, the College may approve requests for alternative conference locations, or alternate means of conducting the conference, such as by telephone, but must impose a special fee of $100 to help defray the additional costs involved.

You are urged, if at all possible, to have the conference at the College, rather than to request special arrangements. If you can coordinate a visit to the College with a vacation period or business trip, the College will make every effort to schedule the Pre-Graduation Conference accordingly.

How can I obtain additional information about the conference?

Consult your academic counselor at the College for information and suggestions on filling out the application form and preparing for the conference. Once the application has been completed and returned, the Office of Test Development and Assessment will review the application and schedule the conference. The Director of Liberal Arts Programs will be happy to answer your questions concerning the College's procedures.

Thomas A. Edison College
Application for Pre-Graduation Conference

NOTE: This application must be submitted to the College at least six weeks prior to the requested interview date.

1. YOUR NAME ______________________ SOC. SEC. NO. ______________
Last First M.I.

2. ADDRESS ____
No. Street City State ZIP Code

3. PHONE NO.: DAYTIME: ______________ EVENING: ______________
Area Code No. Area Code No.

4. SCHEDULE INFORMATION

The College may need considerable leeway on either side of the date you suggest to schedule the conference. Approximately when do you wish to have the conference?

____
Month Year

On the schedule below, cross out with an X any morning, afternoon, or evening period during your usual week when you could NOT be present for the conference.

	MON	TUE	WED	THU	FRI	SAT
MORNING						
AFTERNOON						
EVENING						

Other information or comments related to scheduling of the conference:

____

____

____

5. REQUESTED INTERVIEW LOCATION (please check one)

( ) At Edison College
( ) At another location ($100 special service fee; requires approval of College)

Please give suggested location and reasons for request:

____

____

____

SIGNED ______________________

DATE ______________________

6. AREA CONCENTRATION (check one)

( ) Humanities ( ) Social Sciences ( ) Natural Sciences/Mathematics

A Topic you suggest for discussion in your area of concentration: (Remember, faculty members need not limit themselves to these topics).

A few books or articles you have read on this topic:

Your work, study or leisure activities related to this topic:

B Another topic you suggest for discussion in your area of concentration:

A few books or articles you have read on this topic:

Your work, study or leisure activities related to this topic, if any:

C Another topic you suggest for discussion in your area of concentration:

A few books or articles you have read on this topic:

Your work, study or leisure activities related to this topic, if any:

7. Second area you request for conference (check one)

NOTE: This cannot be the same area you checked in (6).

( ) Humanities ( ) Social Sciences ( ) Natural Sciences/Mathematics

A Topic you suggest for discussion in your area of concentration:

A few books or articles you have read on this topic:

Your work, study or leisure activities related to this topic, if any:

B Second topic you suggest for discussion in this area:

A few books or articles you have read on this topic:

Your work, study or leisure activities related to this topic, if any:

8. Specialization, if you have one: (See pp. 24-25 of the Thomas A. Edison College catalogue).

If you wish to include additional discussion topics, please use the other side of this page.